THE UNIVERSITY
of EDINBURGH

Studying the function of the alternative spliced isoform *Rac1b* in colorectal cancer

THE UNIVERSITY *of* EDINBURGH

Victoria Gudiño

Doctor of Philosophy

The University of Edinburgh

2019

Declaration

I declare that the work from this thesis has been completed by myself, unless acknowledged otherwise. None of this work has been submitted for any other degree or professional qualification except as specified.

Victoria Gudiño

September 2019

Abstract

Colorectal cancer (CRC) is a leading cause of cancer-related deaths, being the secondary most lethal cancer type in UK and the third worldwide. In the recent years, detection methods and therapies have improved overall survival and life quality of patients diagnosed with CRC but it is still a major health problem. There are a few signalling pathways whose alteration are hallmarks for CRC such as the WNT signalling pathway, RAS, TGFβ and the *TP53* gene related signalling. For instance, the *Apc* gene, the core negative regulator of the WNT signalling pathway, is mutated in over 85% of total sporadic CRC cases and it is considered an initiating event for tumour formation. Targeting any of these pathways is the paradigm for conventional anticancer treatments. However, off-target effects or the development of resistance urge for additional treatments. Using multidrug therapies to tackle resilient cells or drugs affecting downstream pathway activation could be promising strategies to overcome treatment failure.

Many molecular processes are altered during cancer evolution in order to favour tumour survival. One that has caught the attention of cancer researchers over the past decades is alternative splicing. Alternative-produced isoforms of proteins have been observed overexpressed in a wide spectrum of tumour types compared to non-tumoral healthy tissues and oncogenic properties have been attributed to many of them. *RAC1B* is an example of an alternative spliced isoform, which was firstly discovered in colorectal cancer cells. Since then, it has been associated with tumour-related functions. Its canonical protein product is RAC1, a small Rho GTPase protein which cycles from active to inactive states and drives cell migration and invadopodia formation, among other functions. Interestingly, Rac1 signalling, but not its expression, has been detected overactivated in tumours, suggesting an oncogenic role for those proteins that promote its GTP-bound state. *RAC1B* is the result of the insertion of 19 amino acids within the switch II domain, which is responsible for the conformational change during its cycling from GTP to GDP. Consequently, RAC1B remains in a GTP bound state and it is considered a constitutive active version of RAC1. Since it was first discovered in breast and colon cancer cells, it has been assigned a tumour-related role. However, little is known about its downstream activated pathways and its function in both homeostasis and cancer.

The goal of the present thesis was to describe RAC1B role in the intestine, either within a neoplastic context or during normal conditions. Besides, I aimed to characterise RAC1B relevance in tumour formation and to elucidate its potential as a therapeutic target.

Initially, *RAC1B* tumour expression was studied through analysis of The Cancer Genome Atlas (TCGA) dataset. Interestingly, it was detected highly expressed in many cancer types such as lung, pancreas, thyroid and gastric-related cancers. Moreover, patients with high levels of *RAC1B* have a worse overall survival and disease-free survival. CRC can be divided into four different subtypes, the CMS subtypes. Curiously, even though *RAC1B* correlated with the canonical/WNT-related subtype, which has one of the best prognosis (CMS2), patients within this group with high *RAC1B* expression had a survival rate comparable to those with the most malignant subtype, CMS4. Therefore, these data indicate that *RAC1B* expression correlates with a worse overall prognosis for CRC patients.

I then studied both its tumour promoter role and its requirement to drive tumorigenesis by modulating its expression in tamoxifen-inducible mouse models, which emulated different stages across tumour evolution. Other than survival and tumorigenesis data, intestinal tissue was used for histologic and molecular analysis, as well as for extraction of intestinal stem cells to culture in a 3D organoid system for following functional analysis. A benign adenocarcinoma model driven by heterozygous loss of *Apc* (APC) and an invasive tumour model with *Apc* and *TP53* co-deletion (APC P53) showed the most interesting results and provided most information about Rac1b-related signalling pathways.

On one side, *Rac1b* depletion in the APC model prolonged mice survival by decreasing tumorigenesis. Phenotypically, these tumours were less proliferative and RNAseq analysis demonstrated a strong WNT-driven phenotype, since most of its target genes were depleted in the co-deleted group. Moreover, a protein interactome analysis by BioID showed a large interactive network, with EGFR and ErbB2 receptors and other RTK and proteins involved in the EGFR signalling pathway. Functional analysis using a colorectal 2D cell line validated its interaction and presented a decreased EGFR activation when *Rac1b* was deleted. Interestingly, APC-derived organoids with *Rac1b* deletion were more sensitive to the treatment with an EGFR inhibitor drug than organoids with wild type *Rac1b* expression, indicating a synergistic effect. Moreover, tumours from the aging cohort also presented a downregulation in two transcriptional targets of the ERK pathway, suggesting that depletion of *Rac1b* was modulating both WNT and EGFR signalling pathways to diminish tumorigenesis.

On the other side, the opposite effect was observed when *Rac1b* was overexpressed. The APC P53 model was used in order to model tumour invasion, although these mice presented a rapid tumour initiation phenotype rather than long-developed invasive tumours. *RAC1B* overexpression drove an increased tumorigenesis in the small intestine which led to a significant decrease in mice survival. Promotion of tumorigenesis was also observed when *RAC1B* was overexpressed in a pre-tumorigenic APC P53 model, which developed significantly more micro lesions or early adenomas compared to the control mice. Intriguingly, these microadenomas presented a marked decrease in the percentage of apoptosis. Transcriptome and functional analysis on the organoids derived from these mice revealed a decreased TGFβ-induced apoptosis, potentially through downregulation of its pro-apoptotic target gene *Bim* and its signal transductor *SMAD4*. Histoscore quantification of endpoint aging tumours stained for SMAD4 antibody demonstrated a significant reduction in immune positivity in the group with *RAC1B* overexpression, pointing to SMAD4 as a potential Rac1b downstream mediator to regulate cell death and TGFβ tumour suppressor role.

Altogether, this thesis provides evidence for a tumour promoter role of Rac1b. Given that its expression positively correlates with a worse patient prognosis, therapeutic targeting of RAC1B in combination with other anticancer treatments could efficiently amplify their effect. Nonetheless, cellular context and tumour stage might vary its downstream function, highlighting the need of studying the tumour mutational landscape prior therapy.

Lay summary

The human body is composed by more than thirty trillion cells, which grow and divide under tightly regulated controls to avoid any cell damage or mutations. When these adverse events occur, our control system detects those altered cells and eliminates them from our body. However, if there is a dysregulation of our check-point controls, mutated cells will accumulate and grow uncontrollably, forming a tumoral mass and giving rise to cancer. Tumours might be seen as parasites, since they will take advantage of normal physiological processes to their own benefit to ultimately spread across different tissues.

Cells contain our genetic information within DNA molecules, which work as an instruction manual for their proper functioning. However, the molecules to conduct their function are not DNA but proteins, which are the result of molecular processes derived from DNA. Different proteins can be originated from a same DNA sequence by the so-called process of splicing. Usually, splicing provides a predominant type of protein but, under certain circumstances, alternative proteins are derived from the same DNA sequence. This process is referred to as alternative splicing and its protein products have been an interesting topic of study in some diseases such as cancer. It is believed that these alternative proteins can provide advantages to the tumours compared to their canonical counterparts. Besides, they are usually more abundant in tumours than in adjacent healthy tissues, providing an interesting therapeutic target.

The present project focuses in the study of an alternative splicing protein product named RAC1B. RAC1B has been predominantly seen in tumour tissues compared to its not alternative protein analogue, RAC1. RAC1 is ubiquitously expressed in all tissues and it has functions related with cell movement and organization of cellular compartments. RAC1 cycles from active to inactive states as a self-regulator of its function. In contrast, RAC1B is almost constantly maintained on its active state, which translates into a continuous functioning, and this lack of regulation suggests a tumoral role. However, little is known about its function in both normal tissue and tumours. In order to gain knowledge about its role, I used mouse models for colorectal cancer (CRC) whose RAC1B expression was either eliminated or expressed at high levels and studied its function and the effects in these mice.

Initially, I made use of a large human tumour data set to study *RAC1B* levels across a wide spectrum of tumour types. I observed that *RAC1B* was mostly expressed in tumours and that colorectal cancer patients with high levels of *RAC1B* had a worse prognosis and lower overall survival. Besides, *RAC1B* expression correlated with other cancer driver genes relevant in a vast range of tumours, supporting the hypothesis of a tumour-related role for RAC1B.

I then made used of CRC mouse models which emulated either a benign colorectal carcinoma or an invasive tumour and I observed two main results. Firstly, when *Rac1b* was deleted in the less aggressive model, mice survival was improved and they developed less tumours compared to mice with normal *Rac1b* levels. This was due to a reduced activity in two cancer-related process (WNT and EGFR signalling), which are usually targeted for anticancer treatments. Interestingly, patients treated with this treatment (inhibitor for EGFR signalling pathway) commonly develop resistances, leading to tumour reappearance and more aggressive secondary disease. However, if the treatment is combined with another drug that reduces *Rac1b* levels, the overall effect is improved. Secondly, when *RAC1B* levels were overexpressed compared to normal levels in an invasive tumour mice model, survival was reduced and tumour formation was increased, worsening mice performance. Following analysis revealed that *RAC1B* was acting in a process responsible for eliminating mutated cells, namely apoptosis, which works as a protector for tumour formation. My results indicate that high *RAC1B* expression suppresses its function, which in turn leads to the survival of mutated cells and tumour formation.

Altogether, this thesis provides evidence of a tumour promoter role for RAC1B. I have elucidated processes and molecular partners RAC1B interacts with, suggesting unknown strategies and points of view to face cancer therapies.

Acknowledgements

I moved to Edinburgh 4 years ago to start my PhD and I cannot be more grateful for that choice, and all the experiences that it has brought to my professional and personal life. Of course, none of this would have been possible without the amazing people that I have been surrounded by over these years.

First of all, I would like to acknowledge my supervisor Kevin Myant for giving me the opportunity to conduct this project in the first instance. Since the first day, you have been a friendly and approachable supervisor and I have been so lucky to learn everything from you first-hand. The door to your (cold) office was always open to hear any crazy not-good-English spoken idea or incomprehensible result, and I have to thank a large part of my current scientific knowledge to our long unscheduled chats. Likewise, I would like to thank the steadily growing Myant group for creating such a nice working environment. Michela, for your enormous patience since the beginning with introducing me to the basics of molecular biology and your cheerful way of working; Caroline, for your infinite help in any Rac1b- or running-related causes. Your personal book of acknowledgements is on the way but, in the meantime, I will just let you know that working side-by-side with you has been essential for juggling all the overlapping experiments and finishing this project. Adam, for your ideas and well-explained protocols, and for always being so keen on offering advice about any topic. Patrizia, for our late-night chats in the office, your help about TGF-β and your interest in Rac1b; it was great having you in the lab and I have really enjoyed working alongside you. Chiara, for being around always when I needed it, albeit your long "commuting" between labs! Antonia, for your cheerfulness and motivation, it was great to have such a positive energy at the end of the project, and Seb, for your short but helpful insights. I would also like to thank Luke Boulter for your close support and for always being available to any questions. To Ed, for our great office talks, to Victor, for your closeness and your scientific tips and to everybody from the W3 floor that helped and contributed to my daily lab work. Furthermore, I would like to express my gratitude to the technicians at the CBS animal unit, without whom the mouse work of this project would not have been possible; especially to Vicky, John, Scott and Linda. Thanks to Alex von Kriegsheim for conducting Mass Spectrometry analyses and to Noor Gammoh for her EGF advice, both of which led to interesting outcomes. I would also like to thank Stuart Aitken for performing our RNAseq analyses, to Craig Nicol for his designing support and to Javier Caceres for agreeing to be my second supervisor. Overall, thanks to everybody at the IGMM for forming part of this experience.

Aside from the scientific aspect, many other people contributed to my thesis without even knowing a thing about Rac1b. En primer lugar, tengo que dar las gracias a todos los que han formado parte de mi pequeña familia edimburguesa. A Georgina i David, per accedir a fer aquell Skype amb zero expectatives de que funcionés però que va canviar-ho tot. Vau ser essencials per fer-me sentir com a casa i la llar que vam crear al 4F1 de Leith és un dels millor records que guardo d'aquests anys, junt amb el Carlos com a hoste especial i essencial a la family. A més a més, em vau introduir a la que és la nostra família d'Edimburg. Paul, Tomás,

Mario, Iván, Giulia, Ate... a todos los betisburgueses, repatriados o aún en tierras escocesas, gracias por demostrar que el frío de Edimburgo es menos frío cuando lo acompañas de buena música, buena comida y mucho amor de churritas. Echo mucho de menos nuestra rutina y estilo de vida en Edimburgo, pero me alegra al ver que se mantiene tanto aquí como en cualquier otra ciudad. Como no, también forman parte de ello Isa y Efrén. Me pareció de las coincidencias más bonitas el conoceros antes de mudarme a otro piso, y fue otro de los acontecimientos más importantes que marcaron esta etapa. Hicimos de nuestro piso en Thirlestane rd un gran lugar para vivir, y no solo por su tamaño. Junto con Alizee y Rachel, you all have supported me over this time with continuous drinks, intensive and deep talks, good Chinese next door take-away food and crowded parties at the flat, and I love you all lots for doing so. Of course, I need to mention Tom and Yan as essential people in keeping me mentally stable along the stressful episodes we all have passed through this time. Our The "whatever is happening at that time" Clubs and our brunches are the best relaxing getaway for any occasion and I miss not having you both in the same city.

Una parte importante de mis agradecimientos también quiero destinarla a mis amigos de Barberá del Vallés. A los xustas, por ser como sois y quereros por ello por mucho que pase el tiempo; y a las wenorras lim (ojalá tuviésemos otro nombre menos embarazoso para poner en una tesis), por estar ahí siempre y hacer que la distancia no marque ninguna diferencia. Os quiero a todas muchísimo, y en concreto me gustaría mencionar a Diana, por las interminables y sorprendentemente frecuentes notas de voz que hemos mantenido durante este tiempo y han hecho tan fácil los momentos de nostalgia. Y a Natalia, por empezar esta aventura de la biología hace 11 años sin tener ni idea que la íbamos a acabar así de bien, y así de juntas, tanto tiempo después. Por nosotras y nuestros yos de 18 años va esta tesis. De la misma manera también se lo quiero agradecer a Marina, Montse y Víctor, por convertir los fines de semana en Barcelona en verdaderos escapes de la rutina escocesa. Y a Carlos, por arriesgarte y hacer tan fácil el vivir en Barcelona.

Esta tesis también quiero dedicarla a mi madre, mi padre y mi hermana; por haberme apoyado siempre en todo lo que he querido emprender y haber sido una ayuda esencial para llevarlo a cabo. Las vueltas a Edimburgo no hubiesen sido las mismas sin los paquetes de comida preparados cada mes y vuestra cara de ilusión al verme aparecer por la puerta. Espero os sintáis orgullosos al verla acabada.

Por último, ya sabes que va por ti. Hace cuatro años nos despedíamos en Barcelona y, si me llegan a explicar nuestra historia, no me la creería. Has llegado a coger más aviones que metros en un mes, te integraste en mi vida en Edimburgo a la misma velocidad que yo y conseguiste hacer que la distancia fuese sólo un pequeño y pasajero inconveniente. Y todo para que ahora esta tesis ya haya acabo, la tuya esté apunto y los vuelos queden como una anécdota lejana. Simplemente, gracias por ser tan Virginio y hacerme la vida tan fácil en todos los países.

Table of contents

DECLARATION **1**

ABSTRACT **2**

LAY SUMMARY **5**

ACKNOWLEDGEMENTS **7**

TABLE OF CONTENTS **9**

LIST OF FIGURES **17**

LIST OF TABLES **17**

ABBREVIATIONS **18**

CHAPTER 1: INTRODUCTION **19**

1.1 COLORECTAL CANCER 19
1.1.1 Statistics 19
1.1.2 Influential factors for CRC onset 20
1.1.2.1 Risk factors 20
1.1.2.2 Protective factors 23
1.1.3 Hereditary CRC 28
1.1.3.1 Lynch syndrome 28
1.1.3.2 Familial adenomatous polyposis 29
1.1.3.3 MUTYH-associated polyposis 30
1.1.3.4 Hamartomatous polyposis syndromes 30
1.1.4 Sporadic CRC 32
1.1.4.1 Adenoma-carcinoma sequence 32
1.1.4.2 The serrated pathway 35
1.1.4.3 New insights onto tumour origin 37
1.1.5 Physiology of the intestine 41
1.1.5.1 Cell populations 42
1.1.5.2 Signalling pathways maintaining intestinal homeostasis 46
1.1.6 Cancer driver genes and their signalling pathways 48
1.1.6.1 The WNT signalling pathway 48
1.1.6.2 RAS/mitogen-activated protein kinase (MAPK) pathway 51
1.1.6.3 The TGF-β signalling pathway 53
1.1.6.4 p53 signalling 54
1.1.7 Stratification of CRC tumours 55
1.1.8 Therapeutic approaches for CRC 58
1.2 RHO GTPASES 60
1.2.1 Rac1 65
1.2.1.1 Actin dynamics 66
1.2.1.2 ROS production 68
1.2.1.3 Signal transduction 68
1.2.2 Rac1 signalling in cancer 71

1.2.2.1 Activating mutations ... 71
1.2.2.2 Degradation defects ... 71
1.2.2.3 Alterations in GEFs ... 72
1.2.2.4 Alterations in GAPs ... 74
1.2.2.5 Alternative splicing ... 74
1.3 ALTERNATIVE SPLICING ... 75
1.3.1 Implications in cancer ... 76
1.3.2 Rac1b ... 78
1.3.2.2 Cancer-associated functions ... 79
1.3.2.1 Alternative splicing regulation ... 85
1.4 ANIMAL MODELS FOR CRC ... 87
1.5 CONCLUDING REMARKS ... 89
1.6 HYPOTHESIS AND AIMS ... 91

CHAPTER 2: MATERIAL AND METHODS ... 92

2.1 MOUSE MODELS ... 92
2.1.1 Modulation of Rac1b expression ... *92*
2.1.1.1 VillinCreER^{T2} Rosa26$^{lsl\text{-}Rac1b/lsl\text{-}Rac1b}$ ... 92
2.1.2 Colorectal cancer mouse models ... 93
2.1.2.1 VillinCreER^{T2} Apc$^{fl/+}$ ... 93
2.1.2.2 VillinCreER^{T2} Apc$^{fl/+}$ TP53$^{fl/fl}$ ... 93
2.1.2.3 Disease surveillance and tumour quantification ... 95
2.2 TISSUE HISTOLOGY ... 95
2.2.1 Fixation, tissue processing and sectioning ... 95
2.2.2 Tissue staining ... 96
2.2.2.1 Haematoxylin & Eosin (H&E) ... 96
2.2.2.2 Periodic Acid Schiff (PAS) ... 96
2.2.2.3 Immunohistochemistry (IHC) ... 97
2.2.2.3.1 Nuclear β-catenin IHC ... 97
2.2.2.4 BaseScope™ ... 98
2.2.2.5 QuPath IHC quantification ... 99
2.3 MOLECULAR BIOLOGY TECHNIQUES ... 99
2.3.1 RNA extraction ... 99
2.3.1.1 Whole tissue ... 99
2.3.1.2 2D cells ... 100
2.3.1.3 Intestinal organoids ... 100
2.3.2 cDNA synthesis ... 101
2.3.3 Real time PCR (qRT-PCR) ... 101
2.3.4 RNA sequencing ... 103
2.3.5 Protein extraction and quantification ... 103
2.3.5.2 2D cells ... 103
2.3.5.3 Intestinal organoids ... 103
2.3.6 Western blotting ... 104
2.3.7 Protein interaction study by BioID ... 105
2.3.7.1 Cloning the BirA* enzyme into a lentiviral vector ... 105
2.3.7.2 Amplification of Rac1 and Rac1b by PCR ... 106

2.3.7.3 Cloning of Rac1/Rac1b into the pLJM1-myc-BioID2 vector 106
2.3.7.4 Streptavidin-biotin pull-down for MS 108
2.4 2D CELL CULTURE 108
2.4.1 Cell line model 108
2.4.2 Routine cell culture maintenance 109
2.4.3 Plasmid transfection 109
2.4.4 Biotin treatment for MS analysis 110
2.4.5 Depletion of Rac1b with morpholino 110
2.4.6 EGF stimulation 111
2.5 3D ORGANOID CULTURE 111
2.5.1 Small intestine epithelial isolation and culture 111
2.5.1.1 Epithelial fractionations 112
2.5.2 Tumour adenoma isolation 112
2.5.3 Routine organoid culture maintenance 112
2.5.4 Clonogenicity assay 113
2.5.5 EGFR inhibition with PD153035 treatment 114
2.5.6 Acute and time course induction with TGF-β treatment 115
2.5.7 TGF-β-induced cell death 115
2.6 TCGA DATASET ANALYSIS 116
2.7 MATERIAL LIST FROM IGMM TECHNICAL SERVICES 117
2.8 STATISTICAL ANALYSIS 118

CHAPTER 3: CHARACTERIZATION OF RAC1B EXPRESSION 119

3.1 LOCALIZATION OF RAC1B IN THE MOUSE INTESTINE 119
3.1.1 Unespecific signal with Rac1b antibody 120
3.1.2 Intestinal epithelial fractionation for Rac1b characterization 122
3.1.3 Designing a BaseScope probe for exon 3b 124
3.1.4 Summary 129
3.2 *RAC1B* EXPRESSION IN HUMAN CRC 130
3.2.1 RAC1B as a tumour-related isoform 130
3.2.2 CMS tumour stratification and correlations based on RAC1B expression 135
3.2.3 RAC1B PSI correlates with genetic alterations in cancer-driver genes 140
3.2.4 RPPA identifies cellular processes associated with RAC1B expression 142
3.2.5 Summary 144
3.3 STUDYING RAC1B ONCOGENIC FUNCTION *IN VIVO* 145
3.3.1 Modelling early adenomas 146
3.3.2 Modelling invasive adenocarcinoma 146
3.4 DISCUSSION 148

CHAPTER 4: RAC1B FUNCTION IN ADENOMA TUMOUR INITIATION 150

4.1 *RAC1B* EXPRESSION IS NOT REQUIRED TO MAINTAIN INTESTINE HOMEOSTASIS 150
4.2 DELETION OF *RAC1B* EXTENDS MICE SURVIVAL THROUGH TWO POTENTIAL MECHANISMS 152
4.2.1 Ageing cohort with Rac1b-deficient mice 152
4.2.2 Histological study of the tumour phenotype 155
4.2.3 The WNT signalling pathway is modulated by Rac1b 158
4.2.3.1 RNAseq analysis 158

4.2.3.2 Acute tamoxifen induction reaffirms WNT-driven phenotype 158
4.2.4 BioID protein interaction study reveals unknown Rac1b-interacting partners 162
4.2.4.1 BioID experiment layout 162
4.2.4.2 Rac1b and Rac1 cloning strategy 163
4.2.4.3 Optimization of BioID technique in CMT93 cell line 163
4.2.4.4 Mass Spectrometry analysis 165
4.2.5 Functional study of Rac1b-EGFR signalling pathway crosstalk 169
4.2.5.1 ERBB family receptors 169
4.2.5.2 Kinetics of CMT93 cells following EGF stimulation 169
4.2.5.3 Exon-3b targeted morpholino efficiently knock-outs Rac1b in CMT93 cells 170
4.2.5.4 Rac1b deletion impairs EGFR phosphorylation and downstream signalling activation 171
4.2.5.5 Organoids from VilCre^{ERT2}Apc$^{fl/fl}$Rac1b$^{fl/fl}$ mice present an EGF-sensitivity phenotype 174
4.2.5.5.1 Impaired cell growth with EGF withdrawal 174
4.2.5.5.2 Increased sensitivity to EGFR inhibitor treatment 174
4.4.3.1 EGFR signalling pathway activation is decreased in Rac1b-deficient mice 176
4.2.6 Discussion 178
4.2.7 Future work 188
4.3 ACUTE ECTOPIC OVEREXPRESSION OF *RAC1B* IN THE MURINE INTESTINE DOES NOT ALTER HOMEOSTASIS 190
4.4 OVEREXPRESSION OF *RAC1B* INCREASES COLONIC TUMORIGENESIS 192
4.4.1 Ageing cohort with Rac1b-overexpressing mice 192
4.4.2 Rac1b-driven colonic tumorigenesis does not appear to be mediated by WNT or EGFR signalling 194
4.4.3 Functional investigation of Rac1b in a time-point model 196
4.4.3.1 Molecular evaluation of the day-31 model 198
4.4.4 Discussion 201
4.4.5 Future work 203
CHAPTER 5: FUNCTION OF RAC1B IN TUMOUR PROGRESSION 205
5.1 *RAC1B* DELETION DOES NOT IMPROVE SURVIVAL WHEN CO-DELETED WITH *APC* AND *TP53* ... 205
5.2 *RAC1B* COOPERATES WITH *TP53* AND *APC* DELETION TO INCREASE INTESTINAL TUMORIGENESIS 209
5.2.1 Rac1b overexpression in an ageing cohort of APC P53 mice 209
5.2.1.1 Tumour histological study 211
5.2.1.2 Rac1b overexpression enhances tumour clonogenicity in vitro 213
5.2.1.3 Intratumour heterogeneity hinders RNAseq analysis 215
5.2.2 Pre-tumorigenic model as a strategy to study tumour initiation 217
5.2.2.1 Early adenomas resemble aging cohort phenotype 217
5.2.3 Rac1b confers resistance to cell death through modulation of TGF-β signalling 221
5.2.3.1 Pre-tumorigenic tissue present different TGF-β activity 221

5.2.3.2 Ex vivo organoid culture to study Rac1b-TGF-β modulation 224
5.2.3.3 SMAD4 as a potential candidate for Rac1b-dependent TGF-β regulation 227
5.3 DISCUSSION 230
5.4 FUTURE WORK 232

CHAPTER 6: DISCUSSION 234

6.1 SUMMARY 234
6.2 RAC1B IS NOT REQUIRED FOR INTESTINAL HOMEOSTASIS 235
6.3 RAC1B AS A PROMOTER OF TUMORIGENESIS 236
6.4 THERAPEUTIC BENEFIT OF TARGETING RAC1B 244
6.6 CONCLUDING REMARKS 246

BIBLIOGRAPHY 247

APPENDIX 1: BASESCOPE CONTROL 271

APPENDIX 2: RNASEQ RESULTS 273

APPENDIX 2.1: RESULTS RNASEQ FROM SECTION 4.2.3.1 273
APPENDIX 2.2: FASTQC FROM RNASEQ EXPERIMENT SECTION 4.2.3.1 275
APPENDIX 2.3: RESULTS RNASEQ FROM SECTION 5.2.1.3 280
APPENDIX 2.4: INTRATUMOUR HETEROGENEITY FROM RNASEQ EXPERIMENT SECTION 5.2.3.1 282
APPENDIX 2.5: FASTQC FROM RNASEQ EXPERIMENT SECTION 5.2.1.3 284

APPENDIX 3: GSEA LIST OF GENES 288

APPENDIX 4: BIOID PROTEIN LIST 290

APPENDIX 5: PAGES ACCESSED 291

List of figures

Chapter 1

Figure 1.1: Colonic tumorigenesis according its mutational route **38**

Figure 1.2: Small and large intestine architecture and their epithelium self-renewal dynamics. **45**

Figure 1.3: The Wnt signalling pathway. **50**

Figure 1.4: The MAPK signalling pathway..... **52**

Figure 1.5: Schematic representation of CRC consensus subtypes. **57**

Figure 1.6: Regulation of Rho GTPase activation..... **62**

Figure 1.7: Actin dynamics orchestrated by Rho GTPases. **67**

Figure 1.8: Schematic representation of genes whose aberrant splicing affects hallmarks of cancer processes..... **77**

Figure 1.8: *RAC1B* transcript and protein activation overview **79**

Figure 1.10: Downstream signalling pathways modulated by Rac1b and their cellular phenotypic translation..... **84**

Chapter 2

Figure 2.1: Mouse models for *Rac1b* expression modulation and CRC..... **94**

Figure 2.2: BaseScope probe for *Rac1b*..... **98**

Figure 2.3: BioID plasmids **107**

Figure 2.4: Schematic representation of morpholino-mediated *Rac1b* knockdown **110**

Chapter 3

Figure 3.1: Rac1b immunohistochemistry of the murine intestine..... **121**

Figure 3.2: Intestinal fractionation experiment characterises a crypt-base location of Rac1b. **123**

Figure 3.3: Visualisation of *Rac1b* transcript in the murine intestine by BaseScope **128**

Figure 3.4: Expression of *RAC1B* increases in human cancers..... **132**

Figure 3.5: Expression of *RAC1B* in CRC correlates with poor prognosis and tumour invasion **134**

Figure 3.6: Expression of *RAC1B* correlates with CMS1 and CMS2 subtypes **137**

Figure 3.7: Featuring traits of CMS1 and CMS2 correlate with RAC1B expression **138**

Figure 3.8: Expression of *RAC1B* subclassifies CMS2 survival outcome **139**

Figure 3.9: Mutational status of cancer-driver genes according to *RAC1B* expression **141**

Figure 3.10: Heatmap correlation of TCGA RPPA and *RAC1B* PSI..... **143**

Figure 3.11: CRC mouse models emulate *Rac1b* expression pattern observed in human cancer **147**

Chapter 4

Figure 4.1: Deletion of *Rac1b* in the normal murine intestine **151**

Figure 4.2: Deletion of *Rac1b* improves mouse survival and reduces intestinal tumorigenesis **154**

Figure 4.3: *Rac1b* deletion reduces tumour proliferation **157**

Figure 4.4: Identification of a WNT signature upon *Rac1b* deletion **161**

Figure 4.5: Set up of BioID experiment in the CMT93 cell line **164**

Figure 4.6: Members of the EGFR signalling pathway interact with Rac1b **167**

Figure 4.7: Deletion of *Rac1b* impairs EGFR phosphorylation and downstream signalling activation **172**

Figure 4.8: *Ex-vivo* and *in vivo* modulation of EGFR signalling by *Rac1b* expression **177**

Figure 4.9: Overexpression of *Rac1b* in the normal intestine **191**

Figure 4.10: *Rac1b* overexpression promotes colonic tumorigenesis..... **193**

Figure 4.11: *Rac1b* promotes colonic tumour proliferation independently of WNT and EGFR signalling activation **195**

Figure 4.12: Short-term overexpression of *RAC1B* does not induce major changes in the murine colon **197**

Figure 4.13: Potential downregulation of *E-cadherin* and *SMAD4* levels in the colon of *RAC1B* overexpressing mice..... **200**

Chapter 5

Figure 5.1: Validation of Rac1b expression modulation in the APC P53 model **207**

Figure 5.2: Deletion of *Rac1b* does not affect tumorigenesis in *Apc*- and *TP53*-deficient mice **208**

Figure 5.3: Overexpression of *RAC1B* promotes tumorigenesis and decreases survival in the APC P53 model **210**

Figure 5.4: Endpoint tumours do not present morphological differences **212**

Figure 5.5: Overexpression of *RAC1B* confers a major clonogenic capacity to APC P53 organoids **214**

Figure 5.6: RNA sequencing analysis of endpoint tumours **216**

Figure 5.7: Characterisation of the day-31 pre-tumorigenic model **219**

Figure 5.8: Rac1b promotes the early formation of microadenomas and reduces their apoptosis **220**

Figure 5.9: TGF-β signalling is decreased in *Rac1b* overexpressing mice **223**

Figure 5.10: Rac1b impairs TGF-β pathway activation **226**

Figure 5.11: *SMAD4* as a potential Rac1b target gene **229**

Chapter 6

Figure 6: Pro-tumorigenic functions of Rac1b **243**

Appendix

Figure appendix 1.1: Positive control for BaseScope staining **271**

Figure appendix 1.2: Negative control for BaseScope staining **272**

List of tables

Chapter 1

Table 1.1: Risk and protective factors for colorectal cancer **27**

Table 1.2: Summary of hereditary disorders in CRC **31**

Chapter 2

Table 2.1: PCR protocol for cDNA and qRT-PCR programs **101**

Table 2.2: Forward and reverse primer sequences used for qRT-PCR **102**

Table 2.3: List of antibodies used for IHC and WB **105**

Table 2.4: PCR protocol for *Rac1* and *Rac1b* amplification **106**

Table 2.5: A-tailing protocol **106**

Table 2.6: Mutational profile CMT93 cell line **108**

Table 2.7: Organoid media conditions and supplemented growth factors..... **113**

Table 2.8: List of solutions and ingredients used in this thesis produced by the IGMM technical services. **117**

Chapter 4

Table 4.1: List of BioID interacting proteins of Rac1/Rac1b **168**

Table 4.2: Rac1b/Rac1 interacting partners related with EGFR signalling **185**

Appendix

Table appendix 2.1: List of significant upregulated and downregulated genes comparing APC and APC RAC1B^{KO} tumours **274**

Table appendix 2.2: List of significant upregulated and downregulated genes comparing APC P53 and APC P53 RAC1B^{OE} tumours **281**

Table appendix 2.3: Individual expression data from APC P53 and APC P53 RAC1B^{OE} tumours..... **283**

Table appendix 3: List of genes within each dataset utilised for GSEA..... **289**

Table appendix 4: Extended BioID protein list..... **290**

Table appendix 5: Pages accessed **291**

Abbreviations

AFAP	Attenuated version form of FAP
APC	Adenomatous polyposis coli
AS	Alternative splicing
BMI	Body mass index
CBC	Crypt base columnar
CC3	Cleaved caspase 3
CD	Crohn disease
CIMP	CpG island methylator phenotype
CIN	Chromosome instability
CMS	Consensus molecular subtypes
COAD	Colon adenocarcinoma
COX2	Cyclooxygenase-2 enzyme
CRC	Colorectal cancer
CRCSC	CRC Subtyping Consortium
DAB	3,3'-diaminobenzidine tetrahydrochloride
DC	Destruction complex
DEG	Differentially expressed genes
DRG	DNA repair genes
Dvl	Dishevelled
ECM	Extracellular matrix
EE	Epithelial extractions
EGFR	Epidermal Growth Factor Receptor
EGFRi	EGFR inhibitor
EMT	Epithelial to mesenchymal transition
ER	Estrogen receptor
FAP	Familial adenomatous polyposis
FPKM	Fragments kilobase exon million mapped
GAP	GTP-activating proteins
GDI	Guanine nucleotide dissociation inhibitor
GEF	Guanine nucleotide exchange factor
GI	Gastrointestinal tract
GSEA	Gene Set Enrichment Analysis
GSK3	Glycogen synthase kinase-3
H&E	Haematoxylin & Eosin
HNPCC	Hereditary non-polyposis CRC
hnRNPs	Heterogeneous nuclear RNPs
HRP	Horseradish peroxidase
IBD	Inflammatory bowel diseases
IGF	Insulin-like growth factor
IHC	Immunohistochemistry
IP	Intraperitoneal injections
ITH	Intra tumour heterogeneity
JNK	c-Jun amino terminal kinases
LI	Large intestine
LOF	Loss of function
LRC	Label-retaining cells
LRP5/6	Lipoprotein receptor-related protein 5 and 6
LS	Lynch syndrome
MAPK	RAS/mitogen-activated protein kinase
MMR	Mismatch repair
MS	Mass Spectrometry
MSI	Microsatellite instability
MSS	Microsatellite stability
NF-κB	Nuclear factor κB
NLS	Nuclear localization sequence
NSAIDs	Non-steroidal anti-inflammatory drugs
PAK	p21-activated protein kinase
PAS	Periodic Acid Schiff
PBR	Polybasic region
PI3K	Phosphoinositide 3-kinase
PIP_3	Phosphatidylinositol-3-4-5-triphosphate
PMO	Phosphorodiamidate morpholino oligomer
PPI	Protein-protein interaction
PSI	Percent Spliced In
ROS	Reactive oxygen species
RTK	Receptor tyrosine kinase
SI	Small intestine
SSA	Sessile serrated adenomas
TA	Transit-amplifying
TCF/LEF	T cell factor/lymphoid enhancer factor
TCGA	The Cancer Genome Atlas
TGFβ	Transforming Growth Factor Beta
UC	Ulcerative colitis
WASP	Wiskott-Aldrich syndrome protein
WB	Western blot
WC	Waist circumference
WT	Wild type

Chapter 1: Introduction

1.1 Colorectal cancer

Colorectal cancer (CRC) is the development of cancer in the colon or rectum. It proceeds with the outgrowth and transformation of benign polyps into adenocarcinoma, which acquire the potential to invade through submucosal layers. If untreated, invasive tumoral cells will eventually penetrate into blood vessels and colonise other tissues by the so-called process of metastasis, which is responsible of the vast majority of CRC-related deaths. Worldwide, CRC is the second most deadly type of cancer although large advances in detection and treatment have improved its overall survival rate. Colorectal cancer refers to cancers developed in both the colon and rectum. Even though anatomic, aetiologic and prognostic differences exist between them, its main traits and characteristics will be next described as a joint cancer type unless stated. The following section will outline the current knowledge in CRC regarding influential factors, classification, cancer driver genes and signalling pathways responsible for its malignancy.

1.1.1 Statistics

Bowel or colorectal cancer (CRC) is cancer originated in the intestine, more frequently developed in the large intestine or colon. It is the third commonest type of cancer worldwide in both, men and women, after lung and breast cancer. However, it is the second leading cause of cancer death only surpassed by lung cancer, accounting for 9.2% of total cancer-derived deaths[1]. In the UK, CRC accounts for the second most common cause of cancer death but it classifies as third when statistics are split according to sex, accounting for 10% of all males and females cancer-related deaths, surpassed by prostate and breast cancer respectively. CRC incidence and mortality are higher in men, since 54% of all CRC deaths are males and 46% are females. Over the last 50 years, CRC mortality has substantially decreased in the UK, reducing by about 44% and by around 14% in the last decade alone. It is one of the cancer types with better treatment response at early diagnosis, since 77% of men and 74% of women survive for one year and about 58% survive for more than five years. The tumoral stage at diagnosis is an important indicator of survival rate, since up to 98% and 90% of patients diagnosed at stage I and II respectively survive, whereas only 40% have good

prognosis when diagnosed at advanced and/or metastatic stages (stage IV)[2]. Still, there has been an improvement of 30% in one-, five-, and ten-years survival over the last fifty years and it is predicted to keep progressing at this rate. However, these numbers might vary largely depending on the country or place of living. For instance in UK, people that live in deprived areas with poor access to health care are diagnosed later and therefore have a worse survival rate (although this has only been demonstrated in men). Worldwide, regions that have less incidence of CRC due to their lifestyle and diet differences, have the same mortality rates to countries with high incidence as a result of limited access to medical controls and the lack of screening programs which prevent a late diagnosis.

Overall, although its incidence has increased over the last years due to environmental risk factors, it is a fairly treatable cancer and improvement in early diagnosis and general awareness have substantially decreased its mortality rate.

1.1.2 Influential factors for CRC onset

A family history of disease (first- and second- degree relatives with CRC) and mutations in CRC-related syndrome genes are the main criteria for an individual being entered into screening detection programs regardless to their age or other risk factors. However, only about 30% of all CRC cases have a hereditary origin and environmental factors highly contribute to sporadic CRC cancer aetiology[3]. Population criteria selection for screening is still under debate and some risk factors that influence CRC incidence do not alter selection parameters (as for example obesity or alcohol consumption). It is estimated that 1 in 15 and 1 in 18 men and women respectively in UK will be diagnosed with CRC on their lifetime, 54% of whose cases could be preventable[4]. Hence, acknowledging its risk and preventive factors could positively modify the chances of developing CRC.

1.1.2.1 Risk factors

Age is the main CRC risk factor with risk steeply rising from 50 years onwards and peaking at its highest incidence risk from 70 to 75 years old. Notably, sex difference is more noticeable at older ages, men being 1.7 times more likely to experience CRC than women

from 65 to 69 years old in UK. Therefore, the majority of screenings programs start at 50-55 years old unless a positive family history is stated[5].

Food digestion and nutrient absorption is carried out in the intestine and the type of food and nutritional properties that pass through the gastrointestinal tract (GI) confers a huge difference and long-term consequences. Red and processed meat consumption has long been associated with CRC risk although studies were never concordant nor definitive. However, in 2015, the World Health Organization (WHO) categorised processed meat within the group 1 (carcinogenic to humans) and red meat into group 2A (probably carcinogenic to humans)[6]. 100g per day of red meat and 50g per day of processed meat increases CRC risk 17% and 18% respectively, mainly due to its cooking (grilled or over-cooked) and processing methodology and the genotoxic metabolites that are generated (H-nitroso-compounds –NOC-, polycyclic aromatic hydrocarbons -PAH- and heterocyclic aromatic amines -HAA-). Association was also observed in other cancers such as pancreas or prostate cancer and relative abundance of these type of meats in some diets contribute to the differential geographic incidence of CRC.

A similar correlation occurs related to alcohol consumption. Moderate (50g/day) and heavy drinking (>50g/day) is associated with CRC risk compared to light or non-drinkers, with a stronger association in men than in women[7]. Indeed, there seems to be a non-linear dose-response to alcohol consumption, with a CRC risk considered for intakes over 30g/day (about 2 drinks)[8]. Carcinogenic mechanisms of action for alcohol include production of both reactive oxygen species (ROS) and acetaldehyde from ethanol oxidation and its role as a solvent to facilitate cellular penetration of other carcinogens (such as tobacco)[9]. Other than CRC, alcohol negatively contributes to other cancer types and malignancies and its socially accepted excessive use is a major unappreciated health problem.

Obesity is another parameter that negatively affects CRC and this is evaluated by the body mass index (BMI), waist circumference (WC) and waist-hip ratio. An excess in any of these three measurements increases the risk of CRC, thought to account for ~11% of all CRC cases in UK compared to normal weight individuals[4]. Meta-analysis studies have classified a BMI over 5kg/m^2 and a WC increased over 10cm as a significant CRC risk[8,10]. Body adiposity causes endocrine abnormalities and metabolic changes, acting as a chronic inflammatory mass and altering pathways such as insulin-like growth factor (IGF) signalling, which

promotes cell growth and inhibits apoptosis[11]. Therefore, lowering overall body adiposity lowers the incidence of most cancer types and improves patient performance during disease.

Interestingly, height is as an unmodifiable CRC risk factor. Compelling investigations have demonstrated a positive correlation with tall people and bowel cancer, with incidence increasing by 5% per every 5 centimetres in height[12]. Potential reasons accounting for this include life factors during childhood such as greater exposure to growth factors and IGF, as well as anatomical reasons (longer intestines and greater number of cells). However, prevention screenings do not consider height parameter as a selection criteria.

Other predisposing factors which do influence screening recommendations are inflammatory bowel diseases (IBD). Ulcerative colitis (UC) and Crohn's disease (CD) accounts for the most common and well-documented intestinal disorders that promote CRC. They consist in the combination of an exacerbated inflammatory response to an altered intestinal microbiota or dysbiosis, accompanied by external insults (stress, smoking, diet, etc) and a susceptible genetic make-up of the host[13]. Twins and familial aggregation studies point to a substantial hereditary load for IBD, since 2%-5% and 5%-8% of UC and CD patients respectively will have IBD family history[14]. Moreover, both diseases could be developed within the same family, suggesting common and shared genetic loci. Their major difference though is their site of origin; CD can develop anywhere in the GI (although it's usually more common in the ileum and colon), whereas UC is specific to the most distal parts of the large intestine. Difference also exist on their location within the intestinal epithelia; while CD inflammation is frequently transmural, UC is confined to the most superficial layers. An early disease onset is displayed in both disorders, usually ranging from 15 to 30 years old[15]. Their symptomatology is chronically discontinuous, with pause and relapse episodes that might mislead the need for eradication. Its cumulative, extended and prolonged inflammation burden is a major risk for CRC development. It accounts for 1-2% of all CRC cases and IBD patients have 70% higher risk of CRC than general population[16]. After 8 to 10 years of initial diagnosis patients are under close and regular surveillance and CRC incidence is 1% higher every year after 20 years of IBD diagnosis. Usually, IBD-related cancers appear at younger age compared to sporadic cancer, about 15-20 years earlier. Their genetics are different to sporadic cancer and carcinoma is preceded by a series of low- and high-grade dysplasia which ultimately culminate in neoplasia, usually multifocal[17]. There is more investigation in UC

compared to CD-associated cancer, although it is believed that both lead to an equivalent cancer risk[18].

Diabetes mellitus type II is an example of non-intestinal disease linked with CRC risk. There is a 30% higher risk of CRC and 70% increased incidence in proximal/right colon cancer in diabetic patients[19]. Mechanistically, the resulted chronic hyperinsulinemia causes elevated insulin and IGF levels which are promoters of tumour proliferation. Controversially, some studies have shown that certain insulin sensitizer anti-diabetic drugs have protective effects for CRC, such as metformin, while insulin secretagogues worsen CRC incidence[20]. Thence, overall CRC risk can be positively and negatively modified depending on the therapy of use.

Finally, economic status is a social factor which influences CRC incidence. Human development index (HDI) accounts for countries whose population have high indexes of life expectancies, education and capital incomes. It is shown that higher HDI countries have more probabilities to develop CRC than countries with low HDI. However, mortality does not follow the same trend, given that better health care facilities and surveillance programs improve survival rate in the richest countries[21]. According to IARC published data from 2018, there is a 72% mortality among all CRC cases developed in low HDI countries, against 45% in high and very high HDI countries, even though there were 40 times more cases in the latter[22]. This disproportional disparity in CRC incidence is independent to the risk factors aforementioned and urge for a global vision when addressing treatment and prevention.

1.1.2.2 Protective factors

Given the high environmental component that enhances CRC onset, there are numerous positive factors which have been demonstrated to be beneficial for high risk individuals and contribute to a healthy lifestyle.

As discussed above, diet plays a major role in CRC incidence. Likewise, evidence exists for some types of foods that counteract other negative stimuli and have cancer preventive properties. Initial interest came after Burkitt's observation of low CRC incidence in African populations, who have a diet based on high-fibre foods[23]. Dietary fibre might lower intestinal diseases and cancer through diverse mechanisms[24]. First, it forms short-chain fatty acids as a result of its fermentation by the intestinal microbiota. One of them is butyrate, which has already been shown to bear anti-proliferative properties. Additionally, it reduces food transit

time and enlarges stool bulk, therefore diminishing the chances for residual mutagens to interact with the lining of the large intestine. It has also been shown to prevent insulin resistance, another risk factor for CRC, and it is a source of bioactive compounds with potential anti-cancer properties such as vitamin C, copper and phenolic compounds. Diverse meta-analysis studies have been conducted to study association with dietary fibre and CRC risk. There is a large amount of heterogeneity and a significant correlation was observed when population was stratified by sex, proving a significant association comparing consumers with high- vs low-fibre diets, regardless of the type of fibre origin[25]. Among the foods with high fibre content are the wholegrains. There is a 17% reduced CRC incidence per 90 grams of wholegrain consumption per day, a meta-analysis from the Continuous Update Project Report (CUP) showed[8], pointing to a protective role for CRC. Another rich source are fruits and non-starchy vegetables. A fruit intake above 300g per day (the equivalent of two average apples) significantly decreases CRC risk, while there is a positive correlation for intakes below 100g per day[8]. Other than the aforementioned fibre properties, fruits are a source of vitamins C and E, as well as for bioactive compounds like folic acid or flavonoids. Non-starchy or resistant starch aliments refers to those foods that escape small bowel digestion and pass straight to the colon for fermentation and consequent formation of short-chain fatty acids. Reduced tumorigenesis has been proven in mice fed with resistant starch, as well as a diminished proliferation at the top colonic crypts[26]. A meta-analysis with eleven studies demonstrated that 100 grams per day of non-starchy vegetables reduced CRC risk by 2% and an inverse correlation was observed for intakes below 100g/day[8]. Aside from dietary fibre advantages, resistant starch vegetables contain a set of anti-tumoral agents whose combination lessen CRC incidence (folic acid, carotenoids, flavonoids and so on).

Other components with CRC preventive properties include foods with vitamin C, D, as well as multivitamin supplements. Vitamin D is mainly known by its role in regulating calcium homeostasis and bone mineral metabolism. However, an anti-carcinogenic function has been gaining prominence after investigating the associated targeted genes of its nuclear receptor VDR, which might modulate colonic mucosa integrity and immunity[27], among other functions. A pooling project conducted by McCullough and colleagues demonstrated a 19% and 27% lower CRC risk with 75nm/L and 100nm/L of circulating vitamin D, respectively[28]. There was no higher improvement at concentrations above 100nm/L, and the opposite outcome was observed for concentrations under 30nm/L, increasing the risk up to 31%. Protective concentrations are over those recommended by IOM (Current Institute of

Medicine) (65,2nm/L), which only take into consideration the amount needed for a proper bone health. Interestingly, they found a better association with women compared to men, probably due to the influence of oestrogen vitamin D activity. Therefore, vitamin D deficiency associates with an increased incidence in cancer, and CRC is the top cancer type with greatest risk according to a report from the WHO[29]. Moreover, it was recently demonstrated that vitamin D levels fall after CRC surgery and this drop predicts a worse patient outcome, suggesting a possible benefit in survival outcome if vitamin D is supplemented after surgery[30]. However, less conclusive information is available regarding the potential protective role for vitamin C or ascorbic acid. It is considered a potent antioxidant and mice experiments have shown an increased uptake in mutated tumoral cells of the oxidized form of vitamin C compared to normal cells, leading to oxidative stress and tumoral cell death[31]. Some clinical trials exploring ascorbic acid therapeutic potential alone or in combination with other chemotherapeutics have shown positive results and apparently, its administration route causes substantial differences on the outcome, with intravenous infusion the most efficient[31,32]. Nevertheless, data regarding its role as a CRC preventive is still limited. Similarly, an inverse correlation exists for multivitamin supplement users compared to non-users, a meta-analysis showed[33]. Overall, additional vitamin supplementation to diet seems to provide beneficial anti-neoplastic results, but a better understanding of its underlying mechanisms would help clarify its contribution and guide therapeutic doses.

Along with vitamin D, calcium has been shown to possess protective functions against CRC. Mechanistically, calcium is able to bind proinflammatory secondary bile acids and ionized fatty acids, as well as reducing cell proliferation and promoting differentiation. The major sources for calcium are dairy products, which also have lactic acid-producing bacteria, casein and lactose that can increase calcium bioavailability[26,34]. Meta-analysis studies from the CUP demonstrate a significant reduction in CRC risk for dairy products (15% decrease per 400g/day), milk (10% for intakes of 200g/day) and dietary calcium (6% per 200mg/day)[8]. However, beneficial role for cheese is limited. Cheese, a rich calcium source, is a double-edge sword due to its high fat content and there is not a significant inverse association with cheese consumption and CRC risk. However, overall, food high in dietary calcium are likely to reduce CRC risk.

A convincing protective factor for CRC is physical activity. Consistently, a number of meta-analysis studies conducted in the last years have shown significant relative risk ratio for

CRC risk in both total and recreational physical activity[35,36]. Numbers range from 20% to 30%, varying with sex or colonic region stratification. Mechanisms associated are a possible reduction in insulin resistance and inflammation and its consequent body weight loss after prolonged high physical activity. However, further studies are needed to validate its protective effect when there is not loss of body weight. Moreover, exercices stimulates colonic motility and reduces transit time, additional factors that contribute to decline CRC incidence[26,36].

Finally, an interesting result is the inverse association with non-steroidal anti-inflammatory agents (NSAIDs), such as aspirin. Through cyclooxygenase-2 enzyme (COX2) inhibition, NSAIDs are potent anti-inflammatory agents. Besides, aspirin has been demonstrated to have colonic specific functions like cell proliferation inhibition by NF-κB and β-catenin signalling interruption and promotion of apoptosis by inducing caspase 8 and 9[37,38]. Specifically in CRC cells, aspirin also inhibits cell proliferation by inducing AMPK/mTOR-dependent autophagy[39]. Given the high expression of COX2 in CRC, attention arose for the use of NSAIDs as anti-cancer or preventive treatment. Meta-analysis trials with low daily aspiring doses (75-300mg) for 5 years revealed a reduction in CRC risk by about 30%[40]. Importantly, at these doses there was a substantial reduction of secondary cardiovascular complications. Recently, it has been shown that aspirin doses administered for anti-cardiovascular and anti-cancer endings should be tailored based on patient's body weight and height for an improved outcome[41]. For CRC chemoprevention, aspirin has been the first agent accepted by the U.S. Preventive Services Task Force (USPSTF) to be routinely used within individuals from 50 to 69 years with no high CRC risk but a specific cardiovascular risk profile[42].

In summary, numerous modifiable factors seem to have, to some extent, a positive and a negative effect on either preventing CRC onset or improving disease course. Expanding general population awareness for a healthy lifestyle against cancer and other diseases is essential, albeit a further elucidation and a comprehensive description of its mechanism of action might help to explode their benefit and potential therapeutic translation (Table 1.1).

		Decreases risk	Increases risk
Strong evidence	**Convincing**	Physical activity	Age Processed meat Alcoholic drinks Body fatness Adult attained height Chronic inflammatory bowel disease (IBD)
	Probable	Wholegrains Foods containing dietary fibre Dairy products Calcium supplements NSAIDs	Red meat Diabetes mellitus type II
Limited evidence	**Limited-suggestive**	Foods containing dietary fibre Fish Vitamin D Multivitamin supplements	Low intake non-starchy vegetables Low intake of fruits
	Limited-no conclusion	Cereals, fatty acid composition, legumes, garlic, non-dairy sources of serum, sugar, coffee, tea, retinol, …	

Table 1.1: Risk and protective factors for colorectal cancer. Adapted from *"The Continuous Update Project Report, 2017"*

1.1.3 Hereditary CRC

Genetics factors are an important factor contributing to CRC risk, believed to account for up to 30% of all cases. However, only 10% of these correspond to known genetic CRC syndromes. The most characteristic trait of hereditary CRC is an early onset, from 25 to 45 years old. Aside from the syndromes described below (Table 1.2), there is a high familial risk if first- or second-degree relatives have developed either cancer or polyposis, which points to a related inheritance whose genetic drivers are mostly still unknown. The major prevention measure to avoid early cancer-related death is a prompt detection and diagnosis, usually by colonoscopy surveillance (easy for polyposis syndromes) or histological/genetic analysis profile of biopsies (non-polyposis)[43]. Below some of the most common bowel cancer syndromes will be described.

1.1.3.1 Lynch syndrome

Also known as hereditary non-polyposis colorectal cancer or HNPCC, Lynch syndrome (LS) is an autosomal dominant condition which causes defects in mismatch repair (MMR) genes and predisposes to cancer in the colon and other organs such as endometrium or brain[44]. Families carrying MMR mutations may develop different primary tumour types segregated within different family members and their detection will facilitate LS diagnosis[45]. Tumours are more commonly developed at the most proximal/right-sided part of the colon and it presents an accelerated carcinogenesis compared to sporadic cancer. Its CRC incidence ranges from 2.5% to 5% of all CRC cases (most frequent inherited CRC type) and has a cumulative lifetime risk of about 40%[46]. LS patients are symptomatic of disease at 30-45 years, slightly later than FAP carriers (Familial Adenomatous Polyposis, described below). Since it belongs to the non-polyposis syndrome group, its diagnosis is more complicated but equally managed by periodic colonoscopy surveillance, yet in some cases subtotal colectomy can be contemplated[47].

Even though it is an autosomal dominant disease, homozygosity is required for an MMR deficient phenotype[48]. This type of malignancy is characterised by a high mutational rate, rendering a mutator phenotype on affected individuals. This in turn facilitates the acquisition of second hits during tumorigenesis. Microsatellite instability (MSI) is also present in LS tumours as a direct consequence of the lack of repair due to MMR mutations. Germline

mutations in at least one of the five MMR genes described comprises LS genetics: *MLH1*, *MSH2*, *MSH6*, *PMS2* and *EpCAM*. Immunohistochemistry to identify loss of any of these gene protein products, especially MLH1 and MSH2, is an easy and rapid method for LS detection. Mutations in *MLH1* and *MSH2* accounts for up to 90% of all LS, while only 10% are due to mutations in *MSH6* and rarely, mutations occur in *PMS2*[45]. *EpCAM* deletion can lead to methylation of the *MSH2* promoter but it is also regarded as a rare event[49].

1.1.3.2 Familial adenomatous polyposis

Within the polyposis syndrome subtype, familial adenomatous polyposis (FAP) is characterised by the emergence of up to thousands of small polyps, predominantly in the most distal region of the large intestine in an autosomal-dominant inheritance fashion[45]. Initial appearance of polyps has an early onset (preteen and adolescents years) and CRC occurs at an average age of 35 years, with an incidence of less than 1% of all CRC cases. Given the large number of polyps, chances that at least one of them accumulates additional mutations and transforms into an adenocarcinoma are high, and therefore CRC penetrance of FAP is up to 100%. Genetically, FAP is the result of truncating germline mutations in the tumour suppressor gene *Adenomatous polyposis coli* (*APC*) and the inactivation of the remaining allele by additional somatic mutation or loss of heterozygosity (LOH). As well as of characterising FAP syndrome, loss of *APC* is the main initiating event in the vast majority of sporadic cancer. It is a tumour suppressor gene due to its negative regulatory role of the WNT signalling pathway and following sections will further describe its role.

Extracolonic manifestations are also seen in FAP patients, which are less frequent and led by different type of mutations in *APC*[48]. Detection of more than 100 polyps confirms FAP diagnosis. An attenuated version form of FAP (AFAP) exists, characterised by developing fewer benign adenomas (hundreds instead of thousands). Consequently, CRC risk is also reduced (up to 70%)[50]. The onset of AFAP is delayed in comparison to the non-mild version (polyposis begins at an average age of 40 years) and its distribution is more proximal rather than left-sided. Diagnosis of AFAP is attributed when less than 100 polyps are detected and its incidence is unknown, although likely ranging from 5% to 10% of all FAP patients[51]. Predominantly, truncating 5' mutations in *APC* occurs, which permits restart of translation and results in a shorter APC protein with constricted functions. Alternatively and less frequently, in-frame deletions due to splice-site mutations derive into hypomorphic *APC*

protein[48]. However, *Apc* mutations partially explain AFAP genetics and the role of modifier genes for Apc function has also been investigated. For instance, there is evidence of some genes located at the 1p35-36 loci and 18q21-23 that attenuates the function of mutated APC by suppressing tumour formation[51]. FAP is clinically managed by periodic colonoscopies according to the tumour burden. Unfortunately, as disease advances, these procedures become more challenging and colectomy surgery is used as CRC prevention strategy. Interestingly, chemoprevention trials using aspirin or natural COX2-inhibitors such as curcumin have shown to be a potential option for AFAP patients whom surgery is avoidable, on top of delaying and preventing new surgeries in FAP patients[52].

1.1.3.3 MUTYH-associated polyposis

Another hereditary condition predisposing to CRC is MUTYH-associated polyposis, which is caused by mutations in the base excision repair (BER) gene *MUTYH*. MYH (*MUTYH* protein) prevents G:C to T:A transversions, and mutations in this gene causes proximal colonic polyposis and an increased risk of CRC[53]. Genetic testing for *MUTYH* is conducted in patients with over 15 polyps and no identified germline mutations in *Apc*[54]. It is estimated that up to 30% patients with 10 to 100 polyps harbour biallelic/homozygous *MUTYH* mutations, and 7% to 29% are associated with individuals with 100 to 1000 polyps[55], which could explain a portion of FAP and AFAP patients without detected *Apc* mutations. Overall, *MUTYH* deficiency predisposes for CRC, with an almost complete penetrance for late-onset disease (over 60 years) and it represents 0.54% of all CRC cases[56].

1.1.3.4 Hamartomatous polyposis syndromes

The vast majority of polyps detected in endoscopies have epithelial origins, which comprises FAP, hyperplastic and adenomatous polyps, among other types. However, polyps can also arise within other intestinal compartments. Hamartomatous polyps are those whose origin is mostly from the mucosal layers and have a disordered tissue growth from in-site cells. Four types of syndromes are characterised by this type of polyp morphology: Peutz-Jeghers syndrome (PJS), familial juvenile polyposis syndrome (JPS), Cowden's disease and Banayan-Ruvalcaba-Riley syndrome. These are rare syndromes although are the most common type among children[57]. They have an autosomal dominant inheritance and only

PTHS and juvenile polyposis are associated with increased CRC incidence. Genetically, PTHS is caused by mutations in the tumour suppressor gene *LKB1*, involved in the mTOR pathway, while juvenile polyposis has silencing mutations in TGFβ-related genes (*SMAD4* and *BMPR1A*)[45,48]. Given the unclassified penetrance of these mutations, their contribution to familial or sporadic CRC is debatable[48]. Similar situation occurs for Cowden's disease, whose malignant transformation is extremely rare but it is likely to enhance sporadic cancer risk given its *PTEN* germline mutation[58].

Disorder	Genes	Inheritance	Onset (years)	Number polyps	CRC lifetime risk (%)
Lynch syndrome	*hMLH1, hMLH2, hMSH6, hPMS2, EpCAM*	A.D.	30-45	non-polyposis	40
Familial adenomatous polyposis (FAP)	*APC*	A.D.	35	100-1000	100
Attenuated FAP (AFAP)	*APC*	A.D.	40	<100	70
MUTYH-associated polyposis	*MUTYH*	A.R.	60	10-1000	80
Peutz-Jeghers syndrome (PTHS)	*LKB1*	A.D.	20	1-100	39
Familial juvenile polyposis	*SMAD4, BMPR1A*	A.D.	42	1-100	39
Cowden's disease	*PTEN*	A.D.	20	-	-
Bannayan-Ruvalcaba-Riley sindrome	*PTEN*	A.D.	20	-	-

Table 1.2: Summary of hereditary disorders in CRC.

1.1.4 Sporadic CRC

CRCs that arise in patients without a family history or genetic predisposition are termed sporadic CRC. Sporadic CRCs occurs mostly in older patients (above 50-55 years), and arise from benign polyp lesions which transform to malign adenocarcinomas after the accumulation of a series of mutations during a long period of time. Up to three quarters of all CRC cases are sporadic tumours whose origin is distinguished by two main routes: the adenoma-carcinoma sequence and the serrated pathway (Figure 1.1).

1.1.4.1 Adenoma-carcinoma sequence

Colorectal cancer was one of the first tumour types characterised at the molecular level, described by Fearon and Vogelstein in 1990[59]. Commonly referred to as the Vogelstein model, it describes the progressive stepwise accumulation of well-defined genetic alterations in initially benign adenomatous polyps up to the formation of malignant and invasive adenocarcinomas.

A preferred order was proposed to exist in the acquisition of oncogenic mutations that leads to tumour formation. Through comparison of mutation frequency on the most mutated chromosomal regions according to tumoral stage (defined by tumour size), they observed a linear sequential pattern of four to five core events which are potentially responsible to drive adenoma progression via clonal expansion of mutated cells[59]. Despite subsequent tumour sequencing projects making major contributions in identifying mutated driver genes, the Vogelstein model is still regarded as the paradigm for understanding sporadic colorectal cancer progression.

Somatic mutations or loss in the *Adenomatous polyposis coli* (*APC*) gene are the first event in the adenoma-carcinoma sequence, giving rise to a hyperproliferative tissue and leading to the initiation of adenomas in about 80% of all CRC cases[60]. *APC* is located on chromosome 5q and germline heterozygous mutations of *APC* are observed in FAP patients. Its product is a large protein of 2843 amino acids, initially thought to mediate its tumour suppressor role through cell adhesion mechanisms but soon after its predominant function was described as a negative regulator of the WNT signalling pathway[61]. In normal conditions, APC ensembles as a complex with the Axin1, Axin2, GSK3β and CK1 proteins to prompt phosphorylation of the core protein β-catenin and direct its targeting to ubiquitin-dependent

proteasome degradation[62]. However, truncated APC is unable to form the degradation complex and therefore β-catenin is stable in the cytoplasm and able to translocate to the nucleus and initiate transcription of its target genes. Aberrant WNT signalling activation is a hallmark for CRC initiation and it will be further described in more detail. An alternative function is attributed to *APC* associated with chromosome segregation and the resulting chromosome instability (CIN). This hypothesis is based on a direct and indirect interaction of APC with the microtubules responsible for a proper assembly of the mitotic spindle[63,64]. Chromosomal segregation errors in still dividing cells promotes tumorigenesis through CIN. An increase in CIN has been observed in *APC*-deficient tumours[65], as well as a result of abnormal WNT signalling activity[66]. Both dual roles, WNT signalling and chromosomal segregation regulation, contributes to its tumours suppressive role and to its function as a "gatekeeper" gene for normal cells in early adenoma formation.

DNA hypomethylation is the second event in the Vogelstein model[59]. DNA methylation is an epigenetic modification that modulates expression of certain genes. The global trend observed in colonic tumours is hypomethylation of key regulatory genes, which also contributes to chromosomal instability[67]. However, in some tumour cells hypermethylation leads to gene silencing, especially on the promoter regions of tumour suppressor genes and DNA mismatch repair (MMR) genes. Highly hypermethylated tumours have been further and separately subclassified as CIMP, CpG island methylator phenotype[68]. They associate with microsatellite instability (MSI) and characterise the alternative pathway for sporadic CRC tumours, the sessile serrated adenomas (SSA), and the previous explained hereditary Lynch tumours.

K-RAS is the most mutated small GTPase protein of the Ras superfamily. Acting as a molecular switch downstream of growth factor receptors, such as Epidermal Growth Factor Receptor (EGFR), it stimulates cell proliferation through the MAPK signalling pathway[69]. Activating mutations, mostly occurring at codon 12, are found in about 40% of all CRCs[70]. Although it is present at a high incidence percentage, it is less common in small dysplastic adenomas (less than 1cm) compared to larger lesions. *K-RAS* mutation is suggested to appear during the transition from early to intermediate adenoma, driving clonal expansion following *APC* deletion[59,71]. However, likely due to its inability to increase the stem cell pool[72], it is not required for adenoma initiation. In recent years, *K-RAS* mutations have also been associated to drive metabolic reprogramming in tumours through overexpression of some glucose

receptors and enhancement of glucose uptake[73,74]. Overall, mutations in *K-RAS* render a proliferative advantage in tumours and contribute to their malignant behaviour.

Loss of heterozygosis (LOH) in regions within chromosomes 17q and 18q are considered the next genetic event[59]. LOH in 18q occurs at a high frequency in all CRC cases (about 70% in all CRC tumours)[75]. While it is infrequent in early or small adenomas, allelic loss of 18q is present in most advanced adenomas and present in virtually all liver metastasis[76]. Originally, the tumour suppressor gene localized in 18q was identified as *Deleted in colorectal cancer* (*DCC*). *DCC* encodes for a transmembrane glycoprotein that mediates the effects of its ligand netrin-1, involved in neuronal cell migration, apoptosis and axon guidance[77]. Even though decreased expression of *DCC* seemed to correlate with 18q LOH in colorectal cancer, as well as in other cancer types, loss-of-function (LOF) experiments in mice failed to consolidate its tumour suppressor role[78]. Instead, a conditional suppressive function for *DCC* depending on its ligand availability rather than a leading tumour inhibition role appears more likely to occur[79]. Further research on genes residing in this locus aside *DCC* identified the SMAD family member 4, *SMAD4*[80]. This is a cytoplasmic transducer of the Transforming Growth Factor Beta (TGFβ) signalling pathway, whose inhibition promotes cell proliferation and resistance to death[81]. 10 to 15% of advanced adenocarcinomas harbour inactivating mutations in *SMAD4*[70], and germline mutations in *SMAD4* are the cause of inherited juvenile polyposis syndrome[45] further supporting the role *SMAD4* as a driver tumour suppressor gene.

Allelic loss on chromosome 17q occurs at a similar rate to 18q LOH and it mainly affects the *TP53* gene[70]. TP53 is a transcription factor for a plethora of target genes and is involved in many cellular processes mainly related in cell cycle check points and cell death[82]. Alterations in *TP53* cause genome instability, promoting tumorigenesis. It is the most mutated gene in all cancers[83] and about 50% of CRC tumours present *TP53* mutations[84]. A higher proportion of *de novo* mutations and loss of *TP53* at adenocarcinomas and invasive foci compared to early adenomas classifies *TP53* as one of the later events during adenoma to carcinoma transition[59,85]. Thence, mutations in *TP53* favour cell division and survival against the cellular stress signals surrounding tumoral cells at late stages.

Finally, additional evidence has led to a further oncogenic pathway incorporated into the original Vogelstein sequence, the PI3K signalling pathway. Phosphoinositide 3-kinases (PI3K) are responsible for the production of phosphatidylinositol-3-4-5-triphosphate (PIP_3), an intracellular second messenger which drives activation of cell growth cues, such as its

canonical and most well-known activation of the pro-survival AKT signalling[86]. The gene that encodes for the catalytic subunit of PI3K, *PIK3CA*, is the second most mutated gene in all cancer types following *TP53*, with an incidence higher than 10%[83]. Activating mutations in *PIK3CA* are seen in about 15% to 25% in CRC, leading to an activation of the pathway independently from its upstream receptor tyrosine kinase (RTK)[70,87]. In addition, the negative regulator of this pathway, the PIP_3 phospholipid phosphatase PTEN, is a well-established tumour suppressor[87]. Mostly inactivated by biallelic mutations, *PTEN* loss is observed in 10% of all CRC tumours, including the rare familial Cowden's disease and in a small percentage of JPS cases[58,88]. The requirement of high WNT activation levels for driving adenoma transformation and its high invasive tumours incidence, both in *PTEN*-deficient mice[89], plus the association between *PIK3CA* and tumour invasion[90], have classified mutations in this pathway to be relevant for the last steps during carcinoma transition.

Accounting for up to 85% of all sporadic CRC, the stepwise sequence progression for carcinoma formation is the cornerstone for the understanding of CRC genetics. Signalling pathways whereby the aforementioned oncogenes and tumour suppressor genes undertake their functions are the top molecular processes investigated and addressed when designing therapeutic strategies. However, further investigations have underscored some drawbacks of this model and suggested new perspectives to consider when studying colonic neoplastic processes, which will be outlined below.

1.1.4.2 The serrated pathway

For a long time, it was believed that the traditional adenoma-carcinoma sequence was the unique pathway for CRC carcinogenesis. However, a subset of polyps different to the conventional adenomas (tubular and tubulovillous adenomas) with a distinguishable histologic and molecular features exist, the so-called sessile serrated polyps. Longacre and Fenoglio-Preiser in 1990 first described lesions with a mixture of features from hyperplastic and adenomatous polyps that had a serrated glandular pattern and different morphological characteristics, which suggested a distinct neoplastic origin[91]. Extensive investigation has been made in these lesions and it was not until 2010 when the WHO officially classified these polyps into three subtypes: the sessile serrated adenoma/polyp (SSA/P), the traditional serrated adenoma (TSA) and the hyperplastic polyps (HPs)[92]. HPs are the commonest lesions among the serrated polyps (80-90%) and are further subdivided depending on their mucin

content[93]. In most of the cases, HPs are benign lesions and difficult to detect in endoscopic surveillance due to their small size, usually less than 5mm. TSA are the rarest serrated lesions (1-6%) with a sessile or pedunculated morphology, normally bigger than 5mm. Both HP and TSA originate in the distal/left colon which accompanies molecular characteristics with less malignant potential[94]. However, SSA/P (15-20%) are the unique serrated lesions located on the proximal colon, their size is larger than 5mm and are considered the main serrated carcinoma precursors[95]. In the last couple of decades, the serrated pathway has been established and well-accepted as an alternative pathway through which carcinomas develop, accounting for up to 30% of all colon cancers.

Cytosine-guanine residues (CpGs) are prone to methylation, which increases risk of cytosine deamination to thymine. As a protective evolutionary mechanism, the genome harbour low frequency of small CpG repeats except for CpG-rich regions in the promoter of some genes, known as CpG islands[68]. Methylation in CpG islands translates into gene silencing of the promoter product, and this can affect tumour suppressor genes. Aberrant or hypermethylation of these regions in a subset of tumours is referred to as CIMP (CpG island methylator phenotype) and classifying them into the serrated route for tumorigenesis[96]. CIMP accounts for up to 30% of all CRC and it is the main characteristic of tumours that originate through the serrated pathway[97]. Depending on the levels of methylation, CIMP is graded as CIMP-L (low), CIMP-H (high) and CIMP-0 (negative), which accompanies a series of specific and distinguishable features. SSA tumours are CIMP-H, develop mostly in the proximal/right colon and are gender-specific (more common in women). CIMP associates with MSI given *hMLH1* hypermethylation. Since both phenotypes cluster in right sided lesions, a proposed mechanism suggests that MSI is a result of CIMP which would eventually lead to carcinogenesis through MMR deficiency[94]. MSI is mostly observed in serrated tumours and CIMP-O and CIMP-L tumours have an MSS phenotype. However, although important, MSI is not an essential mechanism for SSA transformation to carcinomas[98]. Tumours can be classified into a CIMP group according to the hypermethylation status of a previously defined set of genes. Indeed, recent large-scale methylation studies using TCGA data have further subclassified CRC into 5 distinct subgroups, mostly differentiating high CIMP tumours with either *K-RAS* or *BRAF* mutations[99].

MAPK pathway overactivation is the second characteristic trait for serrated tumours. Activating *BRAF* mutations is considered the initiating event of serrated carcinogenesis[100] and

deletions in chromosome arm 5q (*APC*) are very rare. However, SSAs with acquired dysplasia on their malignant transition show WNT activation through missense *APC* mutations or other mutations in the pathway such as *CTNNB1*, both events uncommon in the adenoma-carcinoma sequence[101]. Given the mutual exclusivity with *BRAF* mutations[71], *K-RAS* mutations are found more often in CIMP-L or CIMP-O. Likewise, tumours with MMR-deficiency associate with high *BRAF*-mutation and low *K-RAS*-mutation frequency[98].

Different paths can be undertaken by a serrated lesion during its evolution towards a malignant carcinoma. Still, the molecular traits most commonly seen in these tumours, which constitute the so-called serrated pathway, are MSI in proximal/right lesions with initiating mutations in BRAF and a CpG island methylator phenotype, ultimately leading to carcinoma formation.

1.1.4.3 New insights onto tumour origin

The preconceived way to understand tumorigenesis requires a monoclonal initiating event that through chromosome instability or high mutational rate will facilitate the accumulation of additional driver mutations. Selective pressure within newly acquired mutations and clonal expansion from successful selected subclones will configure the final genetic profile of the transformed tumour. In this regard, the molecular and histological characteristics of colorectal cancer constitute a cornerstone for this principle. However, the study of early and driver mutations that prompt tumour initiation is of great difficulty given the advanced natural origin of human tumour samples. When considering tumour growth as an evolutionary process, one can assume that late sampled tumours would harbour information from early-originated mutations. Hence, evolution of tumour growth and the origin of its intra tumour heterogeneity (ITH) can be assessed at the genomic level. Data obtained from next-generation sequencing (NGS) analysis led to the development of a new model for understanding tumorigenesis (Figure 1.1). Based on the analysis of single tumour glands by multiple genomic techniques, the Big Bang model relies on the assumption that mutations in a tumour do not arise in a linear-sequential manner following clonal expansion, but rather at the initiation of tumour transformation[102]. Presumably, there is an initial transformation with intermixed tumour subclones that will grow as a single expansion. Two types of mutations define these clones: public or clonal, those that will be present in all tumours, and private or subclonal, those that will be pervasive but not dominant. New

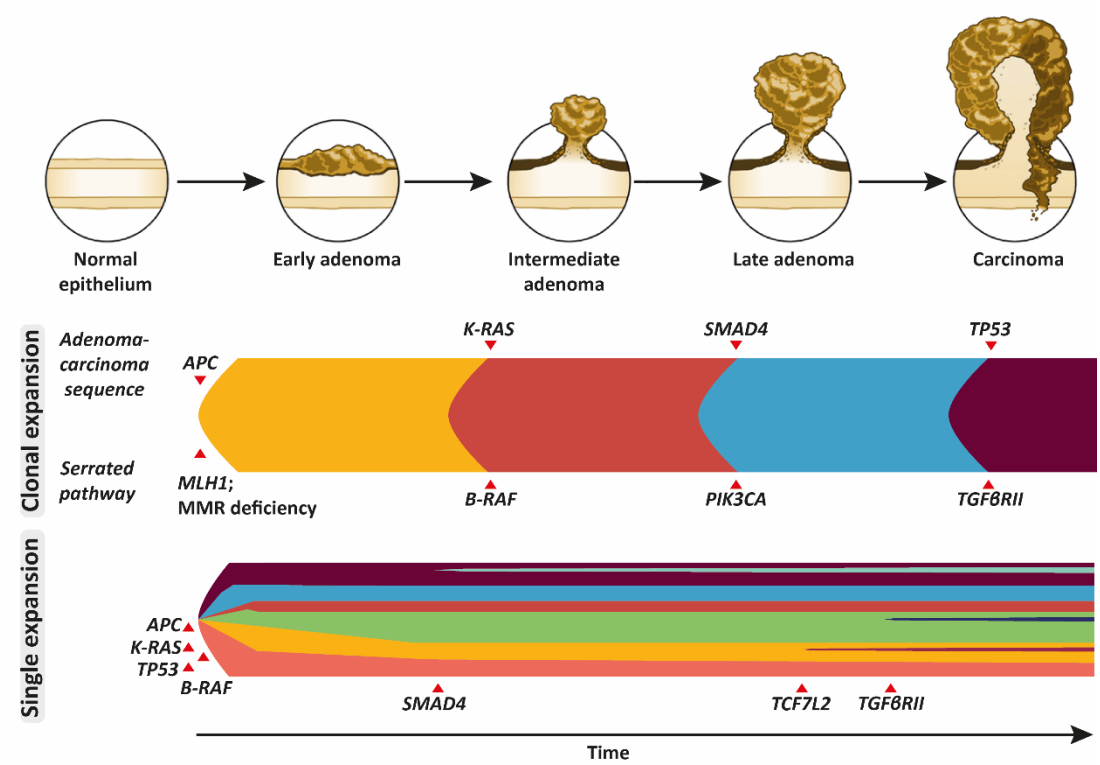


Figure 1.1: Colonic tumorigenesis according its mutational route. According to the clonal expansion model, sequential mutations arise driven by selective pressure. Two tumorigenic routes exist depending on their initiating event (*Apc* loss or *MSI*), the so-called adenoma-carcinoma sequence or the serrated pathway. However, recent advances and data generated by next-generation sequencing (NGS) have allowed the characterisation of a new tumorigenesis theory. The Big Bang or single expansion model postulates that an initial burst of inter-mixed mutated clones arises early on tumorigenesis, which grow without selective sweeps. In both routes, time is the limiting factor.
Mutations presented in the single expansion model are symbolic, based on their percentage of mutations in CRC.
Top figure representing intestinal tumorigenic stages is adapted from *"The molecular hallmarks of the serrated pathway in colorectal cancer (De Palma et al, 2019)"*

private mutations will arise as the tumour progresses due to high mutational rate and replication errors, but these will only be limited to certain subpopulations. According to the model, there are no clonal expansions when new private mutations originate. Even though these mutations might increase tumour fitness, they are not fixed in a selective fashion given that the time of origin prevails over selective pressure[102]. ITH found in most of the tumours supports this model. Previously, it was assumed that heterogeneity was the result of the transition to new driver mutations being fixed. However, the high ITH observed in carcinoma cells does not correlate with a dynamic of linear stepwise selection and selective sweeps. Hence, the model implies that ITH arises early and given the rapid tumour growth rate, fixation of newly formed mutations is a rare event. Besides, it also suggests that aggressive mutations that will provide future advantages on the tumour are already present at early stages, opening a new therapeutic angle. For example, resistance to EGFR inhibitor (EGFRi) treatment is almost unavoidable in virtually all initially responsive patients due to acquired mutations in *K-RAS*[103] and two independent parallel studies found that those mutations were already detectable at low levels in the presumed *K-RAS*^{WT} patients, supporting the Big Bang model for early transformation[104,105]. Moreover, recent exome-sequencing data from paired CRC primary tumours and metastasis demonstrated a low genomic divergence and an early dissemination of metastatic subclones, even when carcinoma is clinically undetectable[106]. Thus, metastatic-associated driver genes are present in early clones and could serve as biomarkers for detection of metastatic-prone tumours. This suggests that understanding ITH in each tumour can be an informative and useful predictor of its future behaviour. The nearly neutral evolution of the Big Bang model for tumorigenesis has been supported by other studies and in cancer types other than colorectal cancer[107,108]. Indeed, studies prior to the Big Bang model already suggested scenarios such as the ones described. For instance, it was generally accepted that some initial intrinsic characteristics of the tumour would anticipate its malignancy such as an increased sensitivity to immune-check blockade in MMR-deficient CRC in contrast to MSS and low-hypermutated tumours[109,110]. Moreover, data generated from the TCGA analysis further supports this model.

Interestingly, exome-sequencing analysis of TCGA data set shows that early-stage tumours present high levels of mutations which are proposed to arise at late stages according to the linear sequence, such as *TP53* or *SMAD4,* indicating that tumour origin might not be monoclonal or that an alternative model to the Vogelstein model may exist. Furthermore, William *et al.* developed a statistical test to validate the model on the TCGA data set[107]. One-

third of TCGA tumours demonstrated a neutral evolution, although the model had a few caveats due to simplistic assumptions needed to be assumed for conducting the test[111]. After improvements on their test, they argue that neutrality is essential for understanding tumour heterogeneity albeit selection remains the main force for adaptation[112].

Even though the Big Bang model has been proven in different settings and it is endorsed by a body of evidence, clonal expansion and selective pressures as driving forces for tumorigenesis should not be discarded. Indeed, the model also suggests that external pressures such as treatment might switch tumour dynamics to a Darwinian evolution to select for treatment-resistant clones[113]. Besides, colorectal cancer mouse models have demonstrated how the addition of certain mutations confers invasive behaviour compared to others and the ability of some mutations themselves to initiate tumorigenesis[114]. Likewise, sophisticated organoid experiments have demonstrated the ability to transform healthy human intestinal epithelium into a carcinoma by stepwise-directed mutagenesis[115,116], and an orthotopic organoid experiment modelled the sequential addition of mutations and the gradual increase of malignant features[117], supporting the contribution of each mutation during CRC progression.

Whether tumorigenesis is preceded by single or clonal expansion, or whether genetic drift or selection pressures are responsible for a subclone predominance is still uncertain. The tissue of origin and initiating genetic event might be highly responsible for the route taken during tumorigenesis. Great efforts are currently being employed to unravel the evolutionary history of tumours, unmasking intricate phylogenetic mutational trees and complex ITH. Despite this valuable information, the requirement of single cell analysis to conduct these studies might be impractical for routine clinical practice. Hence, the expectations of this model in personalising anti-cancer treatment might be challenged by the complex intra tumoral architecture.

1.1.5 Physiology of the intestine

The intestine is the largest structure in the body. In humans it is up to 8 meters in length and has a surface area of over $30m^2$ [118]. It constitutes a physical and biochemical barrier separating the intestinal luminal content from the underlying immune tissue, on top of controlling the intestinal microbiota by segregating commensal bacteria and integrating their microbial signals. Apart from ingestion, it controls all digestive functions, ranging from nutrient absorption to metabolic activities. It consists of the small and large intestine or bowel, the former being further subdivided into three morphological and functional different compartments. The first and smallest section, the duodenum, is in contact with the stomach and harbours the longest villus for nutrient absorption. This is followed by the jejunum and ileum, with villus length decreasing the more distal and closer to the large intestine and mucus secretion increasing for better lubrication of the food bulk towards the colon. The small intestinal wall is composed by four distinguished layers. The mucosa is the innermost layer which secretes digestive enzymes and hormones. It is supported by the connective tissue of the submucosa which connects it to the muscularis layer, in charge of the peristalsis or gut contraction and relaxation movements for digestion. The serous fluid secreted by the serosa, the outermost intestinal layer, lubricates the peristaltic movements from the muscularis[119].

The intestinal villi are finger-like projections from the mesenchyme into the intestinal lumen with the main objective of increasing and maximising the area of nutrient absorption. They are covered by a single cell layer of post-mitotic epithelium and underlined by blood vessels to facilitate nutrient absorption into the organism. Villi are intrinsic of the small intestine, being larger on the most proximal regions. In contrast, the colon is a flat luminal surface lacking villi, given its main function of compacting and excreting stool. Aside from enlarging absorption, villi are exposed to high levels of stress and environmental insults and a quick turnover of their adult intestinal cells protects and decreases the risk of damage exposure to underlying tissue[120]. Each villus is circumscribed at the base by at least six crypts of Lieberkühn[121], which are the niche for the stem and progenitors cells that fuel self-renewal of the epithelium. Stem cells, the so-called crypt base columnar (CBC) cells, give rise to six different epithelial cell types which all except one migrate upwards along the crypt-villus axis.

1.1.5.1 Cell populations

The major characteristic of the intestinal epithelium is its rapid turnover of the adult differentiated cell compartment, which is able to self-renew every 3 to 6 days. This requires a daily generation of over 300 million new epithelial cells[121], which is sustained by the stem cells residing at the crypt bottom. There is only one adult cell lineage whose renewal is slower, the Paneth cells, ranging from 3 to 6 weeks[122]. The six differentiated intestinal cell types are comprised of enterocytes, goblet, enteroendocrine, Tuft, Paneth and M cells (Figure 1.2). Indeed, they can be classified according to whether they have an absorptive (enterocytes and M cells) or secretory (Paneth, goblet, enteroendocrine and Tuft cells) function.

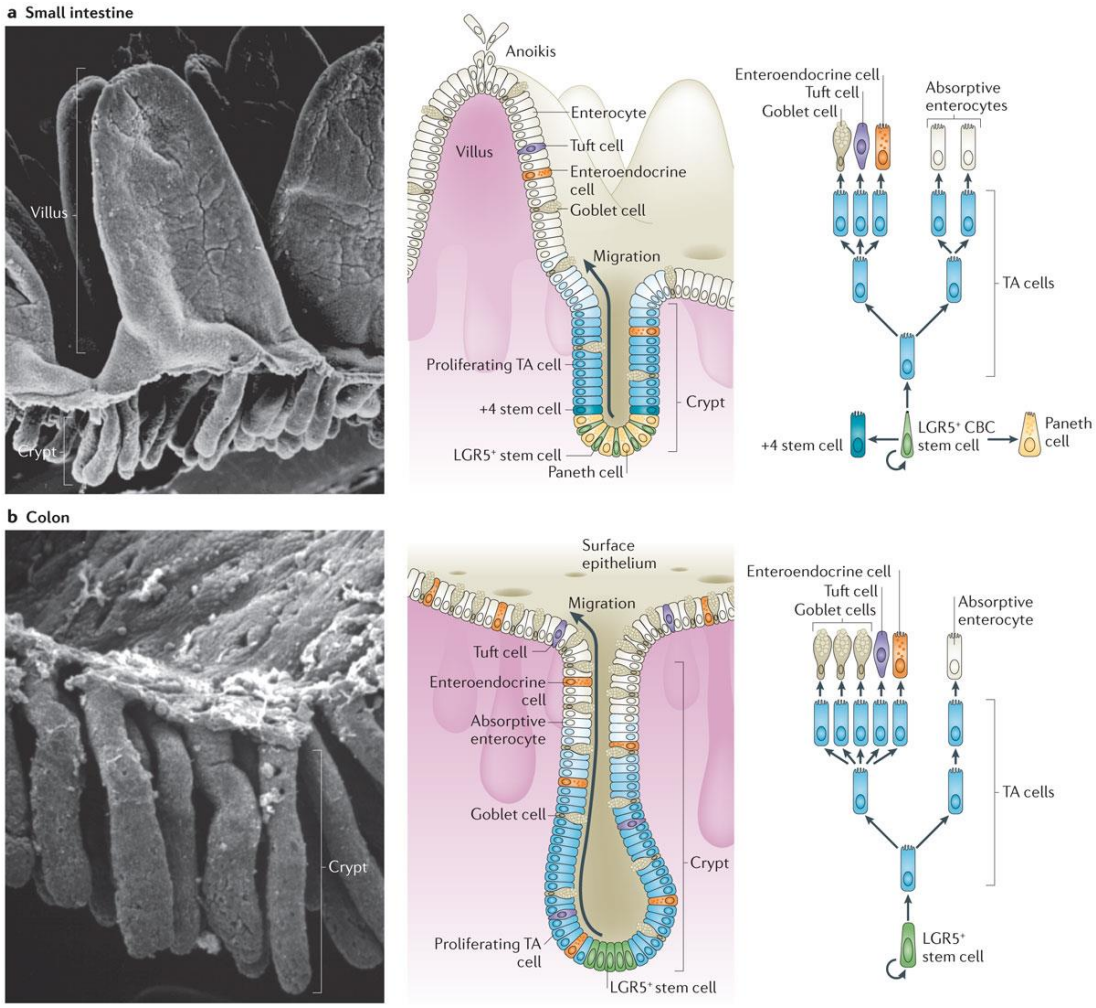
The absorptive lineage is the most abundant in the villus epithelium. Enterocytes process the uptake of nutrients, as well as constituting a cohesive epithelial monolayer for permeability protection. These develop microvilli structures at their apical region, which allow approximately 10- to 20-fold increase of their surface area and establish the so-called brush border, with specialized digestive functions[123]. M cells are localised over the immune cell Peyer's patches and are in charge of internalising luminal antigens from intestinal microbiota and trigger mucosal immunity by the induction of immunoglobulin A (IgA)[124]. Although originally considered secretory cells, Tuft cells may not enter into the full secretory criteria due to the lack of an expressing marker that characterises the rest of secretory cells, *ATOH1*[125]. They are rare intestinal cells that function as a chemosensory cells given their multiple expression of taste receptors signalling members and the presence of a long microvilli at the top. Enteroendocrine cells constitute a small population of the intestinal epithelium, less than 1%, but the largest population of hormone-producing cells in the body. They are able to sense the luminal content and release peptide hormones accordingly, as a coordinated system for gut metabolic functions like insulin secretion or appetite[126]. The morphological cup-like appearance gives the name of goblet cells, which are specialised in mucus secretion. Mucins and trefoil factors are the main components of the intestinal mucus, which serves as a chemical and physical protection from shear stress. These cells are more predominant in the colon and ileum, where mucus properties are beneficial to lubricate the passage of feces[127]. Finally, the Paneth cell linage has a protective function by secretion of hydrolytic enzymes like lysozyme and antimicrobial substances like defensins. The colon is devoid of Paneth cells but instead, other cells positive for $Reg4^{+}$ marker undertake their

analogous function[128]. All six differentiated cell types originate from the intestinal stem cells residing at the crypt bottom and distinct molecular signalling inputs dictate their lineage fate. Transit-amplifying cells are an intermediate cell type between stem and differentiated cells with a high proliferative rate. They reside in the crypt just above the stem cell niche and are capable of differentiating into all functional lineages except Paneth cells, which originate directly from CBC stem cells. Differentiated or mature cells then follow an upwards migration along the crypt-villus axis and cell replacement occurs when "old" cells reach the villus tip by anoikis-dependent cell death. Of note, Paneth cells are the only intestinal differentiated cell type that migrates in a downwards fashion, staying at the crypt bottom interspersing stem cells and sustaining their niche. Given the lack of villus structures in the colon, differentiated cells remain in the top crypt and in the surface epithelium (Figure 1.2).

The stemness criteria of a cell describes the capability to generate multiple lineages and to possess long-term self-renewal capacity. In the intestine, stem cells are constantly proliferating, generating undifferentiated daughter/transit-amplifying cells which divide and differentiate into different cell lineages. It has been largely known that pluripotency of the intestine resides at the crypt bottom, but the lack of advanced genetic tools made further characterization of potential intestinal stem cells challenging. Originally, two independent models attempted to explain their intestine positioning. One of them argued that stem cells resided at the position +4, 0 being the central deepest crypt cell, and that Paneth cells resided from 0 to +3[129]. They demonstrated DNA label-retaining capacity of these cells but progeny linked to them could not be proven. Parallelly, Leblond and Cheng argued that CBC cells represented the intestinal stem cells due to their morphology and cycling activity[130]. However, it was not until twelve years ago that the first stem cell marker for this CBC cells was described using lineage tracing experiments[131]. By the tamoxifen induction of genetically modified mice carrying a LacZ and Cre-EGFP knock-in alleles in the *Lgr5* gene, a WNT signalling target gene, Barker *et al.* were able track cell differentiation from the *Lgr5*$^{+}$ cells. These cells were localised intermingled with the Paneth cells, at positions 0 to +2 and rarely at +4, as it was previously suggested. After five days of induction, the whole crypt-axis was monoclonal and labelled with its expression, containing cells from all lineages. Therefore, *Lgr5* was revealed as a CBC stem cell marker for both small and large intestine. Nonetheless, extensive research has also demonstrated stem cell-like properties for the +4 label-retaining cells (LRC). The current belief for their function argues that they serve as quiescent stem cells which can restore Lgr5^{+} cells and replenish the crypt units following their loss upon

injury[132,133]. Hence, certain circumstances require the LRC pool for proper tissue regeneration. However, it seems that normal tissue homeostasis does not rely on their proliferation, aside from being precursors of Paneth cells[133]. Furthermore, recent investigations have demonstrated that differentiated cells from distinct lineages, including secretory and enterocytes progenitors, can dedifferentiate and repopulate the crypt base, after *Lgr5*$^{+}$ cells loss[134,135]. It is yet unknown whether dedifferentiation ability is marked by niche distance or progenitor stage, although a possible mechanism suggests that contact re-establishment with Paneth cells might provide niche signals to gain stemness[136,137]. Indeed, the stemness of the stem cell zone is also partially maintained by the stem cell niche composed of subepithelial myofibroblasts associated to the crypt, which secrete growth factors and cytokines to sustain stem cell proliferation. It has also been suggested that epigenetic mechanisms might play a key role, given that for instance permissive open chromatin is maintained throughout differentiation and this can facilitate reprogramming[138]. Dedifferentiation plasticity is also observed in other tissues such as the lung or liver[139,140], implying a shared differentiation dynamic with higher complexity to that of unidirectionality.

In summary, the identification of stem cell markers has allowed their thorough study and the characterization of cell growth dynamics along the crypt-villus axis. The stem cell niche and the activity of certain signalling pathways preserve and restrict the stem cell zone at the crypt bottom, which gives rise to all the intestinal cell lineages. Interestingly, adult intestinal cells can cope with the loss of stem cells by self-reprogramming to dedifferentiated states. Despite large contributions, comprehensive mechanisms responsible for this cell plasticity are yet unknown and constitute the current centre of interest in the stem cell field.



<u>Figure 1.2</u>: Small and large intestine architecture and their epithelium self-renewal dynamics. Taken from: "*Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration" (Barker, 2014)*

1.1.5.2 Signalling pathways maintaining intestinal homeostasis

With such complex dynamics to fulfil the rapid turnover for differentiation, a tight signalling network at the stem cell niche and in the transit-amplifying (TA) zone is needed for appropriate functioning and proliferation. Among the most relevant are the WNT, Notch, EGF and BMP signalling pathways, whose ligands are also used for *in vitro* organoid modelling of crypt growth[141].

The WNT signalling pathway and the gene expression profile activated through the β-catenin cascade and TCF4/LEF1 transcription factors have been shown to be the master regulators of stem cells proliferation, as well as for maintaining the constant proliferation of the TA cells. Evidence reside on the disorganization of the stem cell niche and intestinal architecture destruction upon conditional deletion of some of its core elements or targeted genes, as well as on the resultant intestinal hyperplasia and uncontrolled grow when they are activated[142–144]. Binding of Wnt ligands to the Frizzled-LRP5-LRP6 complex and LGR4 and LGR5 receptors leads to dissociation of the β-catenin destruction complex, allowing its nuclear translocation from the cytoplasm, where association with TCF4/LEF1 transcription factors initiates their transcriptomic program (more detail of the pathway will be outlined in the next section). A gradient of WNT signalling activity is observed along the crypt axis, being higher at the bottom and decreasing as it reaches the villus tip. This activity gradient maintains the stem cell zone, which is restricted to the crypt bottom, and delimits high proliferative rate to a confined region. Ligand availability is responsible for this gradient, with WNT3A and R-spondin ligands the best described. WNT3A and other WNT ligands such as WNT2B are produced by Paneth cells and underlying mesenchyme, and it is likely other unknown sources also boost WNT ligands levels[136,145]. Likewise, intestinal R-spondin derives from *PDGFRα*$^{+}$ subepithelial myofibroblast[146]. A gradient of WNT activity is created through the progressive loss of ligand-bound cells across CBC cell divisions, being almost non-existent at the topmost regions[147]. Besides, stem cells express high levels of LGR4 and LGR5 receptors compared to differentiated cells, both specific for R-spondin binding[148]. On top of regulating WNT activity, β-catenin also controls cell positioning through expression modulation of Ephrin ligands and their Eph tyrosine receptors. Receptor expression is restricted to the proliferative intestinal region whilst ligand expression is present in an antagonist pattern, decreasing towards the crypts[149]. This distribution is essential for correct localization of

epithelial cells along the intestinal crypt/villus axis, given the observation that deletion of EphB3 in the murine intestinal epithelium causes Paneth cell mislocalization.

A similar situation as with WNT and R-spondin ligands occurs for EGF ligand. Paneth cells produce EGF and TNFα ligands[136], whose receptor (ERBB1) is highly expressed in CBC cells. The EGF signalling pathway participates in stem cell proliferation but high activity levels can be detrimental and lead to disease. With the aim of controlling a steady state of activity, stem and proliferative cells also express Lrig, its negative regulator of the pathway[150]. EGF is one of the indispensable factors to maintain crypt structures *in vitro*. However, although EGF is essential for their proliferation, it is not required for stem cell identity since stem cells are maintained in organoids upon EGF-depletion, albeit their proliferation is halted and they enter in a reversible quiescent state[151].

Ligands of the BMP family, which activate non-canonical SMAD signalling via the SMAD1/5/8 transcription factor complex, are also important. BMP2 and BMP4 ligands released by intercrypt and intervillus mesenchymal cells promote cell differentiation[152]. In order to protect stem cells from this pro differentiation activity, the stem cell niche expresses high levels of BMP inhibitors such as Gremlin 1, Gremlin 2 and Noggin[153].

Finally, Notch signalling controls CBC cells fate through its lateral inhibition and direct membrane cell contact for activation. When a Notch-activated cell expresses its target gene *HES1*, this in turn inhibits Notch activity on its neighbouring cells through repression of *ATOH1* expression, whereby Notch signalling is also repressed[154]. Notch signalling blocks the secretory lineage differentiation, hence regulating a constant rate of secretory/absorptive differentiated cells through this binary fate.

In summary, generation of the highly diverse specialised intestinal cells and maintenance of the stem cell population is the result of tightly control signalling cascades, whose overactivity or deregulation precedes tissue hyperplasia leading to diseases such as cancer.

1.1.6 Cancer driver genes and their signalling pathways

Transformed cells are characterised by presenting aberrant activation of normal functioning signalling pathways. In the intestine, pathways responsible for tissue homeostasis and cell proliferation are hotspots for deregulation, facilitating unrestricted cell overgrowth throughout the tissue. The TCGA revealed the signalling pathways with most recurrently mutated genes in the majority of sporadic CRCs, which include the WNT, RAS/mitogen-activated protein kinase (MAPK), transforming growth factor beta (TGF-β) and *TP53* signalling[155]. A brief overview on the biochemical characteristic of these pathway will be provided.

1.1.6.1 The WNT signalling pathway

WNT signalling is a conserved pathway with functions in morphogenesis and proliferation during both development and adult tissue homeostasis. There are two functionally distinguishable types with shared receptors but divergent downstream cascades, the canonical or WNT/β-catenin and the non-canonical or planar cell polarity (PCP) pathway. Its activation is driven by WNTs and R-spondin ligands upon binding to Frizzled-LRP5/6 complexes and LGR4/5 receptors.

WNTs are growth factors that promote proliferation through WNT signalling activation. There are nineteen different ligands in humans with likely distinct functions given the phenotypes observed in LOF experiments and their conserved orthologs in other species, albeit their signalling outputs are fairly similar[156]. They are usually co-expressed in the same tissue, as for example in the intestine where six different WNTs are expressed. Differences might reside on their specificity to activate either the canonical or the non-canonical signalling, although their cooperation and cross-reactivity has demonstrated that virtually all WNTs can ultimately activate the β-catenin cascade[157]. In parallel, the other main type of ligand are R-spondins. These are secreted proteins and four different types exist in humans. They promote WNT signalling by both potentiating WNT-driven activation and binding to their specific receptors[148,158]. Interestingly, WNTs and R-spondin ligands have distinct but cooperative functions for WNT-dependent maintenance of $Lgr5^+$ stem cells: R-spondin induces their self-renewal and stem-cell proliferation, whereas WNT ligands induce LGR5, RNF43 and ZNRF3 expression to sustain R-spondin activity[159].

Heterodimers of Frizzled proteins (FZDs) with their co-receptors, low-density lipoprotein receptor-related protein 5 and 6 (LRP5/6), are the primary receptors for WNT ligands, although R-spondins can also signal through them by binding to LRP6. Association of WNT to FZD leads to LRP5/6 dimerization and a conformational change that induces both phosphorylation of the cytoplasmic LRP5/6 tail and recruitment of the positive WNT regulator Dishevelled (DVL) protein on the cytoplasmic region of FZD. DVL stabilises receptor heterodimerization and along with the LRP5/6 phosphorylated sites recruits the scaffolding protein Axin for initiating the signalling cascade[160]. On the other side, LRG4/5 receptors are the main binding partners for R-spondin ligands. LRG5 was firstly discovered as a WNT target gene and as a marker for intestinal stem cells[131]. Both, LGR4 and LGR5 are co-expressed in the intestine but LGR4 is more broadly expressed along the crypt-villus axis. Upon R-spondin binding, LGR4/5 receptors interact with FZD-LRP5/6 complex and trigger WNT signalling, as well as enhancing WNT ligands signals[161]. Alongside receptors, potent membrane negative regulators for WNT signal exists. RNF43 and ZNRF3 are E3 ligases that ubiquitinate cytoplasmic loops of FZDs and LGR4/5 receptors, inducing their proteasomal degradation. They are also WNT target genes, therefore acting as a negative feedback loop. Interestingly, R-spondin can also bind to ZNRF3 and RNF43, releasing WNT receptors from their negative pressure, thereby indirectly inducing WNT activation[162]. Besides, a dual function for DVL was recently discovered, whereby it can mediate ZNRF3/RNF43-dependent ubiquitination of FZD receptors, hence acting as a positive and a negative regulator of the WNT pathway to ensure proper signalling activity[163].

Signal and gene transcription activation is mediated by the activity of a cytoplasmic destruction complex (DC) and the nuclear translocation of the transcriptional coactivator protein β-catenin. The DC is integrated by the tumour suppressor protein APC, the two constitutively active serine-threonine kinases CK1 and GSK3 and the complex scaffolding protein Axin. Under ligand-free conditions, the DC efficiently captures β-catenin by direct binding to Axin. Thence, sequential phosphorylations by the CK1 and GSK3 kinases in its amino terminal region function as docking sites for the E3 ubiquitin ligase β-Trcp, which targets β-catenin for proteasomal degradation and maintains low levels of β-catenin in the cytoplasm. However, upon ligand binding, Axin localises at the membrane, recruited by phosphorylated LRP6 and DVL, which destabilises the DC and dissociates β-Trcp from the complex. Phosphorylated β-catenin uncapable of being ubiquitinated is accumulated in the DC, which saturates the complex and newly synthetized free β-catenin starts to accumulate

in the cytoplasm[164] (Figure 1.3). Previous models suggested that cytoplasmic β-catenin accumulation was due to an inability to form the DC and hence a lack of phosphorylation sites for ubiquitin ligase recognition[165]. However, they were based on overexpression analysis and this updated model was the first based on studying the complex dynamics at endogenous levels. Free cytoplasmic β-catenin is then able to translocate to the nucleus, displace the repressor protein Groucho and engage transcription by binding to the DNA-bound T cell factor/lymphoid enhancer factor (TCF/LEF) family of proteins. This initiates a signalling cascade of genes involved in embryonic and adult stem cell proliferation, whose overactivation leads to the transformation of intestinal epithelial cells[166].

Inactivating mutations in the *APC* gene is the principal mechanism of WNT signalling deregulation and as it was mentioned in the previous section, its loss is present in the vast majority of CRC. Other less frequent mutations within the pathway include the β-catenin encoding gene *CTNNB1*, *AXIN2*, *AMER1* or *TCF7*[155,167]. Alternative mechanisms to activate WNT signalling also include overexpression of WNT ligands[168,169] or silencing of natural WNT antagonist through hypermethylation of their promoter, such as for SFRPs (secreted frizzled-related proteins)[170]. Despite diversity existing as to the precise mechanism via which WNT activation is achieved, it is recognised as a hallmark of CRC[171]. However, its therapeutic targeting is challenging partly due to the redundancy of some pathway members.

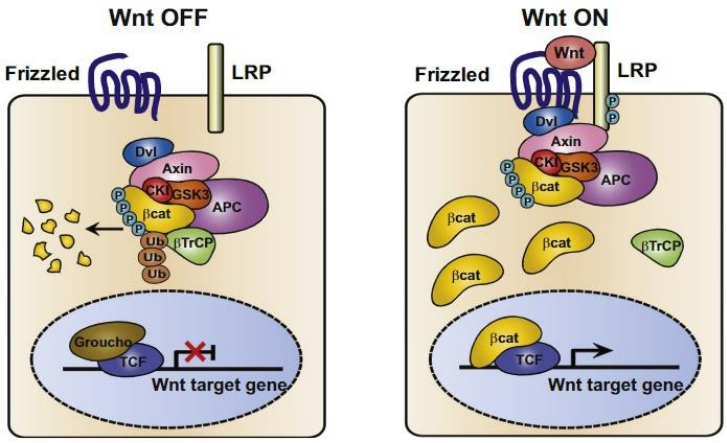


Figure 1.3: The Wnt signalling pathway. In the absence of ligand (left), β-catenin is efficiently degraded through ubiquitination and the pathway remains switch off. However, in the presence of ligand (right), the DC localises at the membrane where the βTrCP cannot bind. The excess of unphosphorylated cytoplasmatic β-catenin translocates to the nucleus and activates its signalling program. Illustration taken from "*Wnt Signaling through Inhibition of β-Catenin Degradation in an Intact Axin1 Complex*" (Li *et al.*, 2012).

1.1.6.2 RAS/mitogen-activated protein kinase (MAPK) pathway

Signalling activated by MAP kinase proteins constitute a sequential phosphorylation cascade that responds to extracellular stimuli such as growth factors, cellular stress and inflammation. Signals are integrated by a set of receptors tyrosine kinases (RTK), which recruit the adaptor protein GRB2 and the RasGEF SOS to activate the Ras small GTPase at the membrane. GTP-bound Ras recruits Raf and initiates the kinase cascade, leading to a range of biological responses such as proliferation, differentiation, apoptosis or inflammation[172]. Four different cascades exist which are commonly referred to by the last kinase member: ERK1/2, JNK, p38 and ERK5. The ERK1/2 (also termed p42/p44) MAPK cascade is the best characterised, especially its Raf-MEK-ERK activation route[173] (Figure 1.4). Activated ERK1/2 translocates to the nucleus for phosphorylation-dependent transactivation of TFs, mainly involved in cell proliferation and differentiation. The JNK cascade is mainly induced in response to cytokines, pathogens, environmental stress or genotoxins. At least fifty JNK substrates have been identified this far, including c-Jun, p53, c-Myc, Bim and MCL1, indicating that proliferation and apoptosis regulation are among its biological functions[174]. Phosphorylated c-Jun can dimerize with JunB, JunD or c-Fos to jointly constitute the transcription factor activator protein 1 (AP-1). Cellular stress can also activate the p38 MAPK kinase, especially stimuli such as UV light, heat or osmotic shock, as well as growth factors like insulin in a cell-dependent manner[175]. Consequently, stress-related responses are activated by this kinase such as Cdc42-induced cell cycle arrest and NF-κB regulation. Besides, it also promotes phosphorylation of other kinases in the cytoplasm like the mitogen-activated protein kinase-activated protein (MAPKAPK). ERK5 is more poorly described but equally important for growth factors and stress response. With a more limited set of substrates, its activation translates in an enhanced c-Jun expression[176].

Epidermal growth factor receptor (EGFR) is the best known RTK that associates extracellular factors with the MAPK signalling pathway. EGFR or ERBB1 is a member of the ERBB family of receptors. Upon ligand binding, EGFR mono or heterodimerises with other members of the family, preferentially ERBB2, and auto-phosphorylates on its tyrosine residues, allowing the association of adaptor proteins like Grb2 and the RasGEF SOS for Ras activation and membrane recruitment of the first MAPKKK. During intestinal homeostasis, EGFR ligands released by the stem cell niche activate proliferative responses through ERK signalling[136]. Indeed, MAPK activity also shows a gradient across the intestinal epithelial cells,

being higher at the crypts and lower at the villi to promote differentiation[177]. Thence, overactivation and expansion of ERK1/2 activity leads to rapid proliferation of undifferentiated cells, promoting cell transformation. An increase in the abundance of EGFR ligands and activating mutations in either EGFR or in the recruited Ras small GTPases confers a cell-autonomous activation of ERK[69]. Three different forms of Ras GTPases exist in humans for RTK downstream activation: K-RAS, N-RAS or B-RAF, which in CRC are mutated at an incidence of 45%, 5% and 10% respectively[69,155]. *K-RAS* and *B-RAF* mutations are mutually exclusive and they associate with distinct tumorigenic routes[71]. All mutations translate into hyperactivity of ERK1/2 signalling but dissecting their mutational profile is of paramount importance for addressing therapeutic strategies, as for instance the benefit of using MEK1/2 inhibitors (MEKi) upon EGFR inhibitors (EGFRi) in *K-RAS/B-RAF* mutated tumours[178].

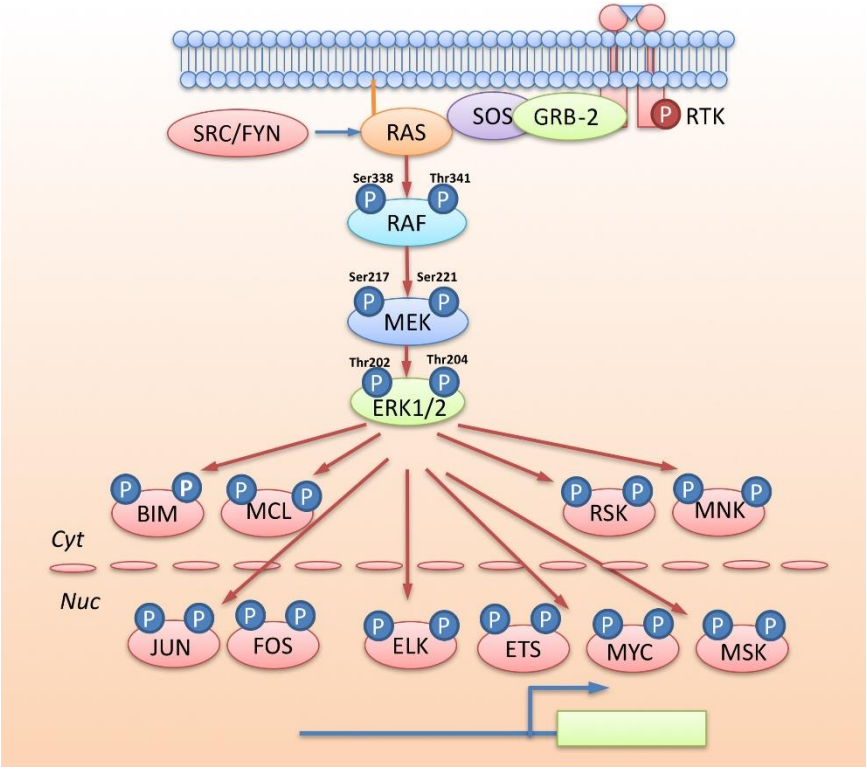


Figure 1.4: The MAPK signalling pathway. The picture illustrates the major downstream EKR1/2 targets, both cytosolic proteins and nuclear transcription factors. Image taken from "Targeting ERK, an Achille's Hell of the MAPK pathway, in cancer therapy" (Liu *et al.*, 2018).

1.1.6.3 The TGF-β signalling pathway

The transforming growth factor-beta (TGF-β) signalling pathway is a superfamily of TGF-β receptors and proteins that are involved in a plethora of cellular processes, such as cell proliferation, apoptosis, differentiation, morphogenesis and migration[179]. It is comprised of 33 different ligands from diverse families (TGF-β, Bone Morphogenic proteins -BMPs-, Activin, and so on) that bind to multimers of TGFβ-1 and -2 serine/threonine kinase receptors (TβRI and TβRII) or BMP receptors (BMPR) and activate either canonical or non-canonical cues. In the canonical pathway, SMAD2/3 are phosphorylated by receptors of the TGF-β family whilst SMAD1/5/8 are activated by BMP receptors. These receptor-activated SMADs then form a trimeric cytoplasmic complex with SMAD4, which facilitates translocation to the nucleus for initiating the transcription of targeted genes[81]. In parallel, the non-canonical pathway includes a SMAD-independent activation of the MAPK signalling, among other pathways such as PI3K or mTOR, through recruitment of Grb2 and SOS on TβRI/II tyrosine phosphorylated sites. Ras and Rho small GTPases and TGF-β-activated kinase 1 (TAK1) associate to the phosphorylated receptors and trigger the ERK1/2, JNK and p38 cascades to ultimately enhance gene transcription[180].

Although TGF-β was first described for its role during embryonic development, current literature mainly focusses on its function in cancer and its switchable dual role. Suppression of the pathway has been observed at early stages of cancer in order to overcome the role of TGF-β signalling in cell growth inhibition and promotion of apoptosis. This can occur by induction of the cyclin-dependent kinase inhibitor p21 or the pro-apoptotic protein Bim[181]. However, in a context and time dependent manner, TGF-β activity changes to a pro-oncogenic function by induction of EMT and inflammation. Cell-autonomous genomic and epigenomic changes in the tumour cell genome might tailor TGF-β anti- or pro-oncogenic functions, such as mutations in *TP53*, loss of *SMAD4* or oncogenic activations in the RAS-MAPK pathway[182]. The TCGA data shows that inactivating mutations in SMAD4 are present in up to 20% of CRCs, along with less frequent mutations in SMAD2/3 and TGF-β and activin receptors[155,183]. However, given its promotion of the EMT process and the highly undifferentiated morphology of tumoral cells when TGF-β is activated, high levels of TGF-β ligands within the tumour or surrounding stroma indicates patient poor prognosis[184]. Consequently, its therapeutic targeting has been hindered by the lack of understanding of its double-edged function and its activity is rather used as a marker for tumour stratification[185].

1.1.6.4 p53 signalling

First cloned in 1983, the p53 protein (encoded by the *TP53* gene) is a transcription factor that regulates cell cycle progression and apoptosis. These functions can also be regulated at the cytoplasmic level given its ability to interact with protein members of the Bcl2 family[186]. p53 activity is inhibited by its interacting protein MDM2, which prevents its function either by promoting p53 ubiquitination or by blocking its trans-activation domain. In normal cells, p53-MDM2 constitute a negative feed-back loop whereby p53 induces MDM2 expression to self-control its function, maintaining low p53 levels[187]. Its activity is also modulated by MDMX, a family member of MDM2 that contributes to MDM2-dependent ubiquitination[188]. Upon DNA damage or cellular stress, MDM2 dissociates and p53 levels increase in order to halt cell cycle and drive apoptosis on damaged cells. Furthermore, regulation of autophagy and induction of senescence are also other mechanisms p53 utilise for suppressing cell growth[189]. Given its critical function, p53 has been designated as the "guardian of the genome" and its proper activity is paramount to keep an organism free from damaged cells[190]. Therefore, it is not surprising that tumours lose its activity to expand clones of transformed tumoral cells.

TP53 is the most mutated gene in all cancer types[83], and the second-most in CRC. It is believed that mutations in *TP53* drives invasion of tumours, although some types of CRC may present p53 loss as the initiating mutational event, such as cancers derived from chronic inflammation[191]. Interestingly, a high percentage of these mutations (74%) are missense mutations, which give rise to a mutant protein that is efficiently expressed in tumour cells and leads to a decrease in wild type p53 activity[192]. Mutant *TP53* is detectable by immunohistochemistry (IHC) in human tumours, which shows a more aggressive morphology and early cancer onset compared to those mutations that cause loss of *TP53* expression[193,194]. Genetically modified mice with either deletion of *TP53* or expression of mutant *TP53* conditionally in the intestine have shown that the latter presents a gain-of-function (GOF) phenotype and induces invasion of intestinal adenocarcinomas[195]. However, a recent study has demonstrated that missense mutations in *TP53* lead to a dominant-negative effect (DNE) rather than GOF activities, equivalent to that of p53 inactivation[196]. In both scenarios, mutant *TP53* is clonally selected over wild type *TP53*. Therefore, restoration of normal p53 is not a feasible therapeutic approach and further elucidation of mutant p53 mechanisms are necessary.

1.1.7 Stratification of CRC tumours

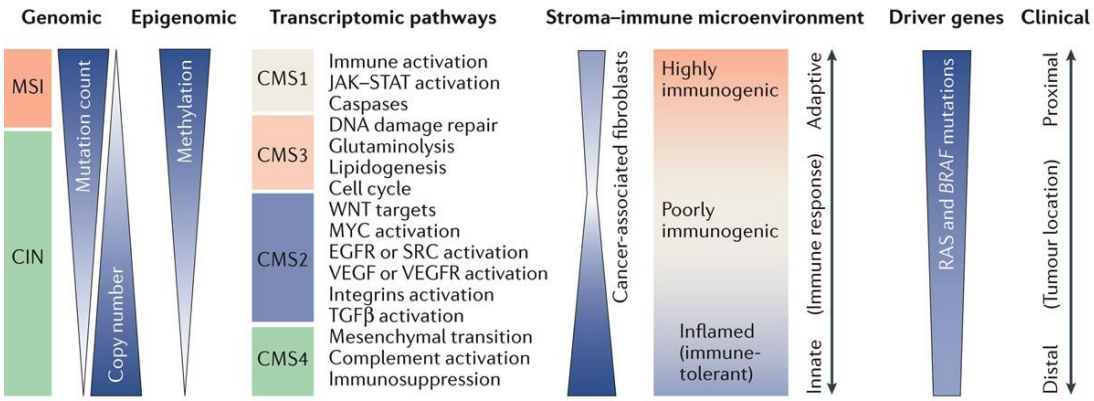
One of the outcomes from the TCGA project was to facilitate the classification of cancer types into subtypes[197,198]. Typically, initial classifications of CRC have been made according to whether tumours displayed microsatellite instability (MSI) or not (MSS) and by their mutational rate[199]. MSI and hypermutated tumours, mostly correlating with *MLH1* silencing, featured a group that represented 15% of all CRCs. Over time and with the acquisition of data from large sequencing programs such as TCGA, it became apparent that tumours harbour further molecular differences which would allow further subclassification. For instance, it was described that within the MSI group, over 37% of these tumours did not have *MLH1* hypermethylation and presented high WNT activity due to *APC* and *K-RAS* mutation, in contrast to the other 63% that showed high frequency of $BRAF^{V600E}$ mutations and low WNT signalling[200]. As the TCGA and other efforts progressed, different groups attempted to describe complex and standardised gene-based classifications but there was a lack of both concordance among all of them and feasible utility for clinical treatment[201–204]. With the aim of using the similarities from each previous approach and to build from them a consistent molecular CRC classification, the CRC Subtyping Consortium was created (CRCSC). It was formed by six different expert groups which had an independent algorithm that previously was used as an approach to subdivide CRC. The CRCSC took eighteen different CRC data sets and normalised their raw data format for these groups to analyse with their algorithm. Each one obtained a number of subgroups, whose association was studied and looked for robust substructures that would suggest clusters between the groups. They then identified four significant clusters, the so-called consensus molecular subtypes (CMSs) (Figure 1.3)[185]. Based on additional molecular data from the cohorts used to establish the CMSs, it was possible to outline the main characteristics for each of the four categories. The CMS1 comprises most of the MSI and hypermutated tumours, it has the highest proportion of *B-RAF* mutations (characteristic of serrated polyps) and an enriched immunological signature. On the other side, CMS2 is the most prevalent and it is considered the canonical CRC subtype. It is enriched with mutations in the WNT/β-catenin signalling pathway and it is mainly MSS. It has the lowest proportion of hypermutated tumours but the highest somatic copy number alteration (SCNA) count, especially in oncogenes. In contrast, CMS3 is considered the metabolic subtype with a high prevalence of *KRAS* mutations and an intermediate MSI phenotype. Last but not least, CMS4 accounts for the subtype with the worst prognosis due to its upregulation in genes implicated in TGF-β and EMT signature. Like

CMS2, it is MSS and infrequently hypermutated and it has a high inflammatory signature as a result of its stromal infiltration. CMS4 is usually diagnosed at advanced stages and presents the worst overall survival and relapse-free survival. CMS1 survival is poor after treatment relapse and CMS2 along with CMS3 shows the best disease-free survival.

Since then, CMS stratification has been implemented in both research and clinic as the "gold-standard" method for characterising tumours based on molecular and prognosis features. However, an evident drawback of analysing whole-tumour expression profile is the joint analysis of tumour and stroma. Evidence exists for a role of the tumour stroma on driving some of the features associated with poor prognosis and invasion, such as the mesenchymal and stem-like traits of the CMS4 and the inflammatory profile of CMS1 and CMS4[205,206]. With the intention of avoiding the stromal bias potentially derived by bulk tumour transcriptome, a new subtyping classification was created. The colorectal cancer intrinsic subtypes (CRIS) consists of the transcriptomic analysis of a large set of patient-derived xenografts using human microarrays, hence differentiating expression of intrinsic only-tumour cells from the stromal gene expression with murine origin[205]. They identified five different subtypes with little overlap with those from the CMS consortium. MSI tumours clustered two major types, CRIS-A and to some extent CRIS-B, which presented an enriched CIMP and hypermutated phenotype compared to the other three subtypes, although MSS samples with a MSI-like phenotype were also present in the CRIS-A group. In contrast, CRIS-B, CRIS-C and CRIS-D were CIN and enriched for WNT activity. Major characteristics associated for each subgroup included *B-RAF* and *K-RAS* mutations with metabolic phenotype for CRIS-A, inflammatory phenotype with high TGF-β signalling and EMT phenotype in the CRIS-B, high EGFR signalling dependency with *K-RAS* wild type tumours in CRIS-C, amplification of *IGF2* and high WNT/stem signalling in CRIS-D, and Paneth-like phenotype with *TP53* mutations in CRIS-D. Hence, major traits from the CMS subtypes were spread into different CRIS groups but the later were absent of stromal content, implying intrinsic tumoral gene profile responsible for driving these phenotypes. Transcriptomes from the xenografts was also validated on their corresponding primary tumours with high fidelity and treatment prediction was identified, such as the significant sensitivity on the CRIS-C group to cetuximab due to their dependence on EGFR signalling and lack of genetic markers of EGFRi-driven resistance. In spite of the novelty of this approach to discern between tumoral and stromal gene expression, bias resultant of the xenotransplantation cannot be completely discarded.

Further clinical validation using this stratification is required and to date it has not had as broad an impact and comprehensive validation as the CMS subclassification.

The CMS subtyping has become highly informative in different settings, albeit current biomarkers used in clinic for tumour prognosis and treatment prediction include as few as MSI and mutational status of *RAS* or *B-RAF*[207]. Incorporation of variables that were not fully integrated in the initial CMS design such as intra-tumour heterogeneity or stromal content, plus other factors that might have biased tumour sample collection like colonic location or tumour treatment might improve the accuracy of the stratification and allow its efficient incorporation for clinical decision making[208,209].



<u>Figure 1.5</u>: Schematic representation of CRC consensus subtypes. Taken from: "*Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer (Dientsmann et al, 2017)*"

1.1.8 Therapeutic approaches for CRC

The therapeutic route taken upon CRC diagnosis varies largely depending on the disease stage at diagnosis. If detection occurs at early stages, surgery is the first option and adjuvant chemotherapy would be administered according to tumoral stage, especially for tumours at stage II and III (with local invasion or with regional lymph-node metastasis)[210]. For metastatic CRC (mCRC), resection of metastasis by surgery is also the best option but unfortunately, the majority of metastasis are unresectable. In that scenario, combination of different chemotherapeutic agents along with the addition of monoclonal antibodies have substantially improved overall survival of mCRC patients up to 30 months in the last 20 years, in comparison to the 12 months expectancy when single chemotherapeutic agent was administered[211]. Hence, a positive change in cancer treatment is observed after the introduction of targeted therapies.

Fluoropyrimidines, especially 5-fluorouracil (5-FU), was initially the standard of care for CRC treatment[212]. Other chemotherapeutic agents also include irinotecan, oxaliplatin and trifluridine/tipiracil. Over time, different regimens of agent combinations emerged, being FOLFOXIRI the one that showed greater anti-tumour activity (triple combination of 5-FU, oxaliplatin and irinotecan)[213]. Indeed, adjuvant chemotherapy with FOLFOXIRI has demonstrated to increase overall survival of patients with stage II to III CRC[214]. However, advances in the molecular staging system have evinced the high levels of intra tumour heterogeneity in CRC and have slowly introduced targeted therapeutic approaches as the paradigm for anti-cancer treatments. Their efficiency relies on accurate patient stratification according to their potential response to therapies, which is achieved by the identification of treatment resistance/sensitivity biomarkers. To date, the genetic alteration most used in clinic for treatment decision making is *K-RAS* mutational status, which allows patient stratification for EGFRi treatment[215]. Indeed, the latest review from the national comprehensive cancer network dictates that all mCRCs patients should be genotyped for mutations in the RAS loci (for *K-RAS* or *N-RAS*) and, if positive, they should not receive EGFRi monoclonal antibodies treatment (such as cetuximab or panitumumab)[207]. Moreover, if they are wild type but present *B-RAF*^{V600E} mutation, EGFRi should be combined with B-RAF inhibitor such as vemurafenib, although its efficiency in CRC is still debated and the clinical trial for this treatment is still ongoing[216]. Another biomarker regularly used is MMR

deficiency, given that stage II tumours with MSI-H do not benefit from 5-FU adjuvant therapy[217].

Efficiency of combined chemotherapeutic agents is increased by additional treatment with cetuximab or panitumumab[218,219] (anti-EGFR antibodies, for *RAS* wild type tumours) or with bevacizumab[220] (an anti-VEGFR antibody, for *RAS*-mutated tumours). As an example, a clinical trial demonstrated that FOLFOXIRI treatment alone on mCRC patients resulted on a median of progression-free survival of 9.7 months, while combination with bevacizumab extended it to 12.1 months[211]. Therefore, the best current adjuvant therapeutic option is the combination of a cocktail of chemotherapeutic agents along with monoclonal antibodies against non-mutated signalling pathways[221]. Unfortunately, this recipe has limited benefit given that a large proportion of treated tumours relapse. Therefore, there is a pressing need to develop more accurate biomarkers to better anticipate tumour response and increase the chances of success. However, as our knowledge in tumour biology expands, so it does its complexity and the level of personalization for the treatment to be efficient. This is accompanied by expensive screenings and high cost techniques which might not be doable to undertake at large scale. In spite of this, the benefit of these approaches might be manifold compared to current chemotherapy and screening standardization and identification of easily detectable new biomarkers will hopefully turn personalised medicine into the future standard practice in oncology.

1.2 Rho GTPases

The Ras superfamily encompasses a large number of small GTPases which influence cell organization and signalling upon stimuli and binding to their associated regulators and effectors. With highly evolutionary conserved orthologs in both unicellular and multicellular organisms, Ras superfamily proteins are divided into 5 main subfamilies: Ras, Rho, Arf/Sar, Ran and Rab[222]. All of them share a common G structural domain which renders all members of the superfamily highly similar but still allows binding and affinity to specific protein effectors. This domain consists of a set of conserved G boxes (G1-G5) that directs their binding and hydrolysis of GTP/GDP, acting as molecular switches for a plethora of intracellular pathways. When bound to GTP, they are in an active state with high affinity for their downstream targets, whereas GTP hydrolysis confers an inactive GDP-bound state. This conformational switch is facilitated by two loop regions, switch I (residues 30 to 38) and switch II (residues 59-67) regions, whose reorientation modifies their affinity for their effectors[223]. There is also an effector core domain located within the switch I region (in residues 32-40) which has proven critical for their direct effector association and specificity[224]. Although simplistic, this process is tightly regulated by guanine nucleotide exchange factors (GEFs), which catalyse GDP to GTP exchange in the membrane, by GTP-activating proteins (GAPs), which accelerates their low intrinsic GTPase hydrolysis activity, and by guanine nucleotide dissociation inhibitors (GDIs), which sequester GDP-bound GTPases in the cytoplasm to avoid binding to GEFs. Besides, additional spatiotemporal regulation exists at the post-translational level in some cellular contexts, that modulates their activation as well as their cellular targeting[223,225]. For example, membrane targeting of Ras proteins is conducted by lipid post-modification on a C-terminal CAAX motif, shared in the majority of Ras and Rho family members. Prenylation (farnesylation or geranylgeranylation) on the cysteine residues of their CAAX box and/or palmitoylation, the covalent binding of fatty acids, allows stable membrane positioning and Ras activation through GTP binding[226]. Indeed, GDI proteins, specific negative regulators of the Rho and Rab subfamily, carry out their cytoplasmic block through masking of the prenyl modification[227]. However, some proteins might undergo phosphorylation for cytosol translocation or ubiquitylation such as RhoA[228].

In humans, the Ras superfamily comprises over 150 different members. Although their biochemical function as GTPases is mostly shared across all members, differences on their G

boxes sequences and specific functions allow their further functional classification in the aforementioned five branches. Besides, GDP/GTP-bound regulation through post-translational modification and through binding to GEFs/GAPs are also specific for each subfamily. Phylogenetically, recent research points to the Arf subgroup as the root for the rest of the superfamily, contrary to the initial belief of Ras proteins as founding members[222]. Given the primordial role of Arf GTPases as membrane trafficking regulators, it underlies the evolutionary sophisticated function of Ras as modulators of cellular processes. The most functionally diverse group is the Ras subfamily, firstly known as viral oncogenes and now described with manifold roles in cell proliferation, survival, differentiation and gene expression control[229]. Rab and Arf are mainly involved in membrane and vesicular transport of proteins, whereby they can also transduce signalling cues[230,231]. The Ran subfamily encompasses the most abundant small GTPases in the cell and are regarded as a subgroup of the Rab family, becoming specialised in nuclear transport of RNA and proteins[222], in addition they have a role in coordinating nuclear envelope and spindle apparatus assembly during mitosis[232]. Rho GTPases are closely related to the Ras subfamily and are characterised by their role in actin cytoskeleton organization, through which they can modulate cell polarity, morphology, migration and adhesion[233]. In turn, they also modulate a wide range of cell proliferation and survival processes via direct or indirect regulation of targeted gene expression. The best characterised members of this family are Rac1, RhoA and Cdc42, and given the central interest of this thesis, a comprehensive description in the Rho subfamily followed by an in-depth depiction of Rac1 will be provided.

The Rho (Ras homologous) family constitutes a group of 20 small GTPases divided into 8 subfamilies[234]. Rho was the first and serendipitously discovered in 1985, whose encoding gene was isolated from the sea snail Aplysia[235]. Short after, in 1989, Rac1 and Rac2 were identified as substrates for the botulinum toxin C3 ADP-ribosyltransferase[236] and Cdc42 was elucidated in 1990 after cloning the formerly known *G25K* gene[237]. All of them presented high homology to the already described *Ras* gene (previous $p21^{ras}$) but shared higher similarity between them, hence were considered to belong to the same subfamily. Almost all of the Rho proteins function as canonical GTPases cycling from active GTP-bound to inactive GDP-bound states, working as molecular switches for signalling modulation. However, eight members of this family are regarded as “atypical” Rho proteins. These remain in constant activation due to key amino acid substitutions that either prevent GAP binding or cause high

intrinsic nucleotide exchange activity, relying on their expression levels or post-translational modifications for their function regulation[233].

GEFs and GAPs are the major regulators for downstream Rho functions, along with GDIs (Figure 1.4). To date, there are over 80 RhoGEFs and RhoGAPs and 3 different RhoGDIs described. GEFs subclassify according to two unrelated carrying domains: Dbl-homology (DH), the most abundant with over 70 members in humans, TIAM1 being an example[238], or Dock homology region (DHR) domains, a family with 11 members specifics for the Rac and Cdc42 subfamily. In contrast, a unique catalytic BCR homology (BH) domain exists for the described RhoGAPs[239] which is unrelated to the other GAPs from the Ras superfamily. Given the very slow intrinsic hydrolysis activity of small GTPases, GAP binding is essential for tuning the signalling cascade to the cellular signal dynamics. Rho family functions are pleiotropic and the large number of different GEFs and GAPS are partially responsible for their broad functionality and signal complexity.

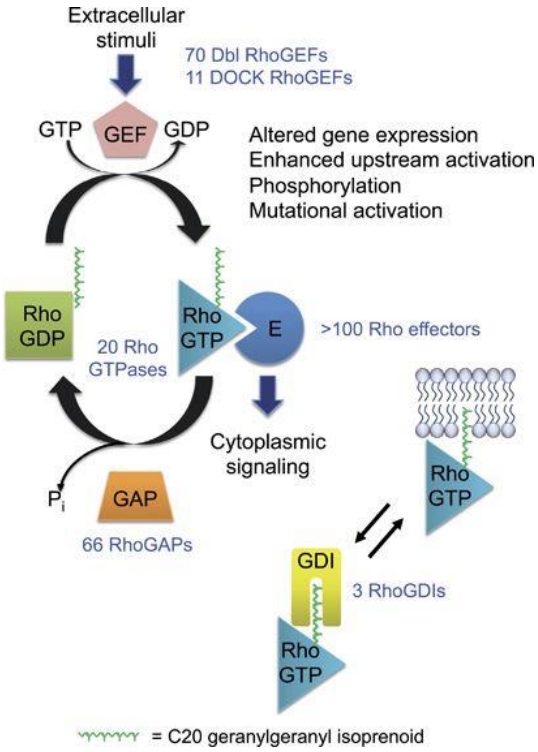


Figure 1.6: Regulation of Rho GTPase activation. Most of the Rho GTPase proteins cycle from GDP-bound (inactive) to GTP-bound (active) states. This process is regulated by GEFs and GAPs proteins, as well as by post-translational modifications that allow their membrane targeting. GDI act as negative regulators by retaining RhoGTPase in the cytoplasm. Figure taken from *"Rho guanine nucleotide exchange factors: regulators of Rho GTPase activity in development and disease" (Cook et al, 2014).*

Rho proteins are activated through stimuli transduced by growth factor receptors and integrins, among others[240]. Intensive research has been made in the last years to draw a better and more complete picture of the downstream functions of each Rho subfamily. Most of the current knowledge has been gained through analysis of cellular changes upon microinjection studies of dominant negative forms of RhoA, Rac or Cdc42, as well as with *in vivo* and *in vitro* experiments with constitutive active or dominant negative forms of each GTPase. Regarding actin dynamics and remodelling, the distinguishable traits for the most well-studied Rho protein subfamily are stress fibre formation and focal adhesion (FA) assembly for RhoA, the promotion of lamellipodium and membrane ruffling for Rac1 and the filopodium formation for Cdc42[234]. Hitherto, up to fifty different effectors for these three Rho proteins have been identified, embracing a broad set of signalling pathways including cell-cycle progression, axonal guidance and chemotaxis, the alteration of which is of relevance for many diseases.

Three highly homologous isoforms, with about 88% of shared sequence, comprise the Rho proteins: RhoA, RhoB and RhoC. All of these are capable of inducing stress-fibre and FA formation, which consist of bundles of contractile actin filaments emanating from extracellular matrix and cell membrane assemblies[241]. In comparison to Rac and Cdc42, Rho act at the rear part of the cells to promote detachment for cell locomotion[242]. At the molecular level, Rho activates the serine/threonine kinases ROCK1 and ROCK2 for myosin light-chain (MLC) phosphorylation and Myosin II activation, which interact with actin filaments from FAs, thereby enhancing contractility and modulating adhesion cell attachment traction (Figure 1.5)[243]. Likewise, Rho interacts with protein kinase C-related kinases (PRKs) and with the formin protein mDia, the latter in charge of promoting actin polymerization as well as of microtubule stability regulation[244]. Given that the three Rho proteins have a highly conserved core effector domain, it was presumed that they would also share most of their effectors; explaining why a large proportion of functional studies were initially conducted using RhoA as a model. However, different cellular location due to distinct post-translational modifications indicate specific targeted proteins and effectors, as it has been demonstrated in the last decade. While RhoA and RhoC are predominantly in the cytosolic compartment, RhoB localizes at the plasma membrane and at the endosomes[245]. This has been translated into substantial differences on their roles in cancer. Functional *in vivo* experiments and human tumour expression analyses have established RhoA and RhoC as consolidated oncogenes, with RhoC having a role as a metastatic driver[244,246]. On the

contrary, RhoB is mainly considered a tumour suppressor gene since it inhibits cell growth and promotes apoptosis. However, some stimuli might induce its expression such as DNA damage or growth factors, indicating that its role in tumours might be context dependent. In fact, it has been shown to regulate endosome trafficking of growth factor receptors such as EGFR, hampering their recycling and therefore, indirectly modulating their signal positively[247,248]. Nevertheless, in most scenarios, RhoB acts as a tumour suppressor gene[249] and its downregulation negatively correlates with patient prognosis.

Initially discovered as an essential factor for yeast budding and cell polarity, Cdc42 (sharing 80% homology with the Rho proteins) was subsequently cloned in mammalian tissue[237,250]. Functionally, Cdc42 triggers the formation of the actin-based structures filopodia, thin and finger-like actin extensions at the front of the cell important for chemotaxis and motility purposes[251]. Like Rho which also assemblies FA for stress fibre association, Cdc42, as well as Rac1, stimulates the assembly of analogous multimolecular focal complexes. These contain proteins such as vinculin, paxillin and focal adhesion kinase (FAK) and associate filopodia and lamellipodia to the plasma membrane[251]. Actin polymerization of Cdc42 is performed by the Arp2/3 complex, which will nucleate new actin filaments and cross-link them into Y-shaped branches[252]. Activation of the complex is mediated by the Cdc42 effector Wiskott-Aldrich syndrome protein (WASP) and its homologous N-WASP, although other effectors can also promote actin polymerization in an Arp2/3-indepent route such as the formin mDia (Figure 1.5)[253].

In brain development, filopodium structures, along with lamellipodia, are essential components during synapsis formation and axon extensions. Mice with conditional knock-out of *Cdc42* in the brain had a reduced number of axons and are lethal when born, likely due to a decreased activity in its effector cofilin and consequent inhibition of filopodial dynamics[254]. In other cell types, loss of *Cdc42* might be compensated by additional Rho GTPases that can also induce filopodia such as Rif, TCL or Wrch-1[255]. Besides, crosstalk between Cdc42 and Rac1 also influences their activities, such as the indirect lamellipodia formation by Rac1-dependent Cdc42 activation and partial filopodia inhibition by reduction of Rac1 activity[251,255].

Filopodium are involved in cell migration, whose aberrant over-activation is observed in pathologic conditions such as cancer[256]. Besides, cancer driver pathways have been demonstrated to be influenced by Cdc42 activity. Cdc42 also establishes cell fate

determination by interacting with the Par6-Par3-atypical protein kinase C (aPKC) complex and abolition of Cdc42 causes polarity defects[257]. Likewise, EGFR signalling is also affected when Cool-1, a Cdc42-specific GEF, is abnormally phosphorylated and transduces into virtually constant activation of Cdc42, mediating its dual block for receptor degradation and ubiquitination[258].

1.2.1 Rac1

There are three different Rac isoforms, without taking into consideration the atypical non GDP/GTP-cycling RhoG[233]. Rac1 is the most highly studied Rac protein, being fairly ubiquitously expressed across all tissues. In contrast, Rac2 is specific to hematopoietic cells and Rac3 to the nervous system. Such a distinctive pattern of expression suggests large differences on their biological functions which will probably be not redundant. Evidence of the wide variety of functions carried out by Rac1 compared to Rac2 and Rac3 arose from genetically modified mice. Whereas whole-body *Rac1*-knockout mice are embryonic lethal, *Rac2*- and *Rac3*-knock-out mice are viable and only present cell-type specific defects[259–261]. *Rac2*-deficient mice showed a substantial diminution of actin content on their neutrophils, which impaired their motility and orientation although other functions such as superoxide production were compensated by the co-expression of *Rac1* in these cells[261]. On the contrary, *Rac3*-knockout mice presented no obvious anatomic differences to their controls albeit subtle defects on their behaviour that improved their motor skills[259].

Subcellular compartmentalisation of Rac1 influence its activation[262]. When inactive, Rac1 is a cytosolic protein forming a GDP-bound complex with a GDI chaperone in a 1:1 ratio[263]. Geranylgeranylation and prenylation modifications and a preceding polybasic region (PBR) to the CAAX motif are responsible for Rac1 targeting to the plasma membrane, where binding to GEFs will confer its active GTP-bound state. Curiously, the same PBR harbours a nuclear localization sequence (NLS) that allows Rac1 translocation to the nucleus[264,265]. Over 30 different effectors interact with GTP-Rac1 bound protein, influencing a myriad of signalling pathways whose functions will be briefly summarised.

1.2.1.1 Actin dynamics

Actin polymerization driven by Rac1 leads to the formation of membrane ruffling and lamellipodia, essential events for cell motility[266]. Membrane ruffles are actin rich membrane protrusions that usually precede lamellipodia formation, which are flat and short actin filaments arranged as a dense array localised at the leading edge of a motile cell[252]. Lamellipodia-driven migration occurs when protrusions are restricted to one side of the plasma membrane[267]. Integrin participation in the focal complexes during lamellipodia formation induces Rac1 activation as a response of their feedback loop, which perpetuates Rac1 activation and makes migration more efficient[268]. As for filopodia, Arp2/3 complex also participates in lamellipodia extension but this is activated through the WAVE (WASP-family verprolin-homologous protein) complex, since Rac1 does not directly interact with WASP[269]. Alternatively, Rac1 can also induce lamellipodia, either by removing capping proteins at the barbed ends, hence allowing further polymerization at the blocked end filaments[156], by activating mDia2 formin, or by modulating cofilin phosphorylation through its effectors p21-activated protein kinase (PAK) and LIM kinase (LIMK) (Figure 1.5)[269]. Actin remodelling is carried out by the combination and spatiotemporal coordination of all three (Rho, Cdc42 and Rac1) activated GTPases. Their resultant actin structures conduct locomotion in migrating cells, either in normal or transformed cells[270]. The clearest example is the negative feedback regulation between RhoA and Rac1 in both their activity and location within the cells, where opposite positioning orchestrates synchronised cell movement (Figure 1.5)[270]. Their coordinated activity might explain why cells can migrate albeit lacking Arp2/3 or WAVE complexes[271,272].

Rapid cell adhesion dynamics are in close relationship with their migration process, given that detachment and rapid reorganization of cell-cell contacts are essential for proper cell integrity. Upon contact between migratory cells, the establishment of a strong cell adhesion involves cytoskeleton remodelling and rapid lamellipodia formation at the contact sites, resulting in active Rac1 accumulation[273,274]. Halting of cell movement and the commencement of adhesion also requires a steady balance in lamellipodium length. Myosin activity partially regulates this process by both blocking filament elongation at the back of the lamellipodia and organizing extracellular matrix (ECM) and cytoskeleton association[275]. Rac1, along with FA-integrin, modulates myosin activity through PAK- and PKC-dependent phosphorylation of myosin II[276]. Fluorescence resonance energy transfer (FRET) experiments

have revealed that activated Rac1 preferentially localizes transiently at *de novo* cell-cell contact sites[277]. Certain plasma membrane compositions such as lipid rafts and phospholipid-enriched zones attract GEF accumulation and therefore promotes activation of Rac1[278–280]. Continuous dynamic movement of activated Rac1 around the cell edge plus actin-myosin complex contractibility at cell contact sites serves as a current model for understanding Rac1 modulation of cell adhesion[281].

Actin remodelling is also necessary for vesicular trafficking, macropinocytosis and phagocytosis. Intracellular uptake of bacteria, antigen presentation complexes or dead cells requires actin polymerization at focalised membrane regions where these particles will be internalised. Membrane reorganization initiates with membrane ruffling at hot spot sites and Rac1 activation peaks before membrane closure following vesicle (or macropinosome) formation[282]. WAVE complexes and PAK1 activation is necessary for Rac1-mediated macropinocytosis[283]. Constant Rac1 activation would impair vesicle formation so Rac1 activity is tightly controlled, thus governing efficient internalization[284].

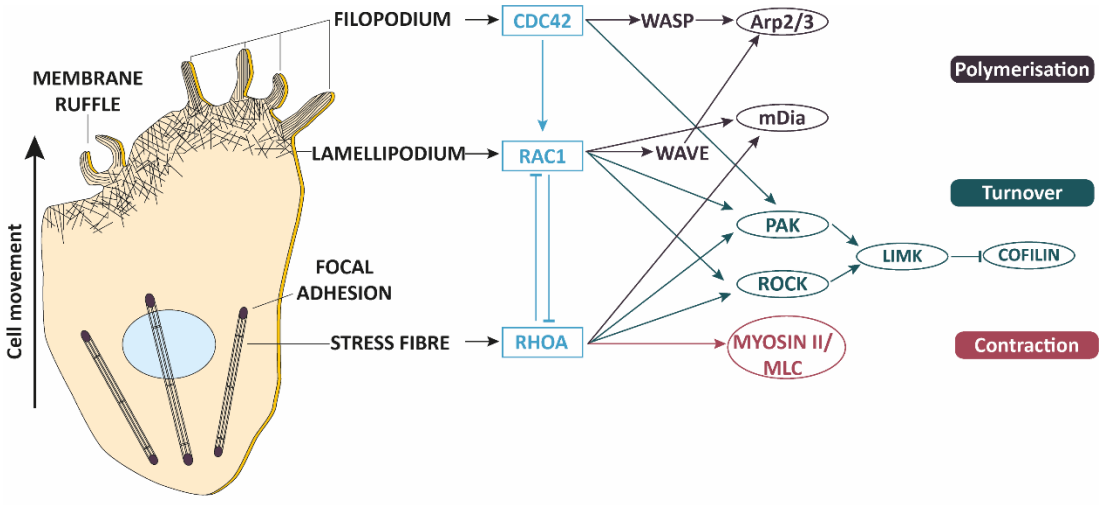


Figure 1.7: Actin dynamics orchestrated by Rho GTPases. Stress fibres localise at the rear front of the cell and regulate cell contraction and mobility. In contrast, lamellipodia and filopodia localise at the leading edge and the combination of these three structures induces forward cell movement. Filopodia is controlled by Cdc42, Rac1 promotes lamellipodia and membrane ruffle formation and RhoA induces stress fibres and focal adhesion. Their protein effectors and crosstalk during actin polymerisation and turnover are schematised in the figure.

1.2.1.2 ROS production

Following Rac1 characterization as a C3 transferase substrate[236], its first functional description was the activation of the NADPH oxidase complex through its target p67phox[285]. NADPH reduces oxygen on superoxide molecules, which can spontaneously form hydrogen peroxide and eventually generate reactive oxygen species (ROS)[269]. Rac1-generated ROS is currently a well-accepted source of ROS, which contributes to a wide range of functions. Among the most extensively described are bacterial killing on macrophages and neutrophils and regulation of senescence through DNA damage[286,287]. Additionally, ROS production mediates Rho GTPases crosstalk, given that Rac1-derived ROS indirectly activates p190Rho-GAP for Rho downregulation[288]. Cell survival is also affected by ROS production in a Rac1-Tiam1-dependent manner through ERK pathway activation[289]. Similarly, ROS levels control stem cell production in a tissue-specific manner[290,291] and it has been shown than in the intestine, generation of ROS drives abnormal stem cell proliferation outside their niche, a mechanism tightly associated with neoplastic processes[292].

1.2.1.3 Signal transduction

Rho GTPases can regulate cellular events beyond processes related to cytoskeleton remodelling. In particular, Rac1 modulation of gene expression has largely been implicated in cell proliferation and death responses, both in homeostatic contexts and human diseases.

The activation of at least three different MAPK pathways have been shown to be influenced by GTPase activation. On one side, it was initially discovered in 1995 that Rac1, along with Cdc42, could stimulate JNK activation[293]. c-Jun amino terminal kinases (JNK) are activated upon cellular stress and regulate the activity of the transcription factor *c-Jun* and its partner *c-Fos*, which upon dimerization form the transcription factor activator protein 1 (AP1). Other subfamilies of MAPK are the p38 and ERK1/2 MAP kinases. All three follow distinct pathways of activation downstream of Cdc42, Ras and Rac1 activation, the latter performing its action though its effector Pak[294–296]. Signal transduction cascades driven by MAPKs are an essential network for cell proliferation as well as for cell survival, given that they can also promote apoptosis in a context-dependent manner. Hence, GTPases such as Rac1 have additional functions as integrators of external signals and gene expression transducers.

Signalling downstream of PI3K also seems to be influenced by Rac1 activation. The substrate of this kinase, PIP_2, was initially studied as an associated partner of Rac1 because of its ability to interact and promote actin polymerization[297]. Rac1 can activate PI3K signalling either by direct interaction or by indirect association of its Rho-GEFs to PIP_2[298]. Activation of PI3K leads to the production of the second messenger PIP_3 and the recruitment of the kinase Akt, which promotes survival responses such as inhibition of caspases and increased expression of anti-apoptotic members like Bcl2 proteins[299]. Thence, Rac1-PI3K interaction activates Akt upstream[300]. Likewise, Akt can also mediate activation of Rac1 through PI3K/AKT-dependent Tiam1 phosphorylation and activation[301]. An alternative method through which Rac1 could potentiate Akt signalling is by the mammalian target of rapamycin mTOR. The principal function of mTOR is to integrate catabolism and anabolism signals but it also targets Akt for phosphorylation, potentiating its full activity[299]. Recently, it has been demonstrated that Rac1 directly interacts with mTOR and functions as a critical mediator of its activity[302], which could potentially suggest the coordination of both signals as a dual outcome of Rac1 activation[303].

Rac1 is required for cell cycle progression by stimulating G_1 phase progression and DNA synthesis through induction of Cyclin D1[304]. Likewise, it might also influence cell cycle by its nuclear entry in G_2, although the mechanism is yet unknown[265]. Several Rho-GTPase effectors have described functions contributing to mitotic entry, such as phosphoinositide-dependent protein kinase-1 (PDK1) and PAKs 1-3[305,306]. Rac1 seems to be the main PAK activator given its localisation at the centrosomes during G_2, whereby it regulates centrosome separation and mitotic entry[307].

Cell survival and proliferation signals stimulated by Rac1 are also driven by induction of the transcriptional activity of nuclear factor κB (NF-κB). NF-κB is a ubiquitously expressed transcription factor (TF) composed by hetero- and homodimers of the Rel family proteins, which remain sequestered in the cytoplasm through binding with IκBα[308]. TF nuclear translocation and interaction with kB responsive elements triggers a wide profile of gene transcription, with a predominant component of immune- and proliferation-related responses[308]. Initial mechanisms through which Rac1 regulates nuclear of RelA translocation included inhibition of its negative regulator IκBα and Rac1-derived ROS activation[309,310]. Rac1-mediated IκBα degradation is undertaken in the membrane ruffles along with the E3-ligase complex SCF[311]. It can also activate the NF-κB signalling cascade in cooperation with PI3K/Akt

pathway, which form a signalling complex with activated Rac1 and promotes TF activation in response of Toll-like receptor 2 (TLR2) phosphorylation[312]. Recently, a novel role of Rac1 has been described, involving the signal transducer and activator of transcription 3 (STAT3). By direct binding with STAT3, Rac1 facilitates the formation of a STAT3-NF-κB complex and its nuclear targeting, with STAT3 potentiating NF-κB signalling[313].

The role of Rac1 in regulating the WNT signalling pathway is more controversial. A body of literature has widely accepted Rac1 as a nuclear translocator of β-catenin, which can therefore initiate the transcription of genes mediated by TCF7/LEF1 TFs[314,315]. Firstly, it was shown that Rac1 would translocate β-catenin to the nucleus through JNK2-dependent phosphorylation[316]. Additionally, Rac1 activity was suggested as an indispensable mediator of colon tumorigenesis for efficient β-catenin nuclear translocation in cooperation with *K-RAS* mutation upon *APC* loss[317]. The latter hypothesis is controversial since tumours without *K-RAS* mutation harbour nuclear β-catenin and β-catenin functions as a canonical read-out of WNT signalling activation regardless of *K-RAS* mutational state or Rac1 activation. Interestingly, work from Myant *et al.* demonstrated nuclear accumulation of β-catenin in the intestine of *Rac1*-deficient mice, indicating that Rac1 operates downstream of canonical WNT activation to enhance its signal[292]. Likewise, Rac1 also plays an active role in the non-canonical WNT signalling pathway, which orchestrates cell polarity through activation of effectors other than β-catenin-TCF/LEF[318–320].

Altogether, the broad variability of interacting effectors and GEFs regulators as its specific subcellular compartmentalization renders Rac1 with a plethora of roles which encompass a wide variety of biological processes. Elucidation of its functions by means of loss- and gain-of-function experiments provides clear evidence that deregulation of either its expression or activation can lead to physiological aberrations such as cancer.

1.2.2 Rac1 signalling in cancer

The first indication that Rac1 could drive cellular transformation came from early studies that showed Rac1 is required for *K-RAS*-dependent neoplasia[321,322]. Later on, validation of this data was provided in a conditional *K-RAS*-driven model of lung cancer[323]. The mechanisms through which Rac1 can be deregulated in cancer are fivefold. As with many oncogenes, it can be activated via activating mutations or by defects on its degradation pathway, hence increasing its expression and/or activity. Likewise, GEFs activating mutations, GAPs downregulation or alternatively spliced isoforms might enhance Rac1 activity.

1.2.2.1 Activating mutations

Some point mutations in *RAC1* have been described in human cancers in recent years that correlate with poor prognosis. Melanoma has the highest occurrence of gain-of-function mutations in *RAC1*, with P29S mutation found in 4% to 9% of tumours in two large-scale whole-exome sequencing (WES) studies and in about 5% in the TCGA study[183,324,325]. $RAC1^{P29S}$ has an enhanced binding to its effectors PAK1 and MLK3, an increased migratory ERK-dependent activity and a better induction of membrane ruffling[325]. Interestingly, this hotspot mutation confers resistance to RAF inhibitors in melanoma cells, suggesting it may be a potential predictive stratification biomarker for melanoma patients[326]. This mutation was also identified in head, neck and endometrial cancers, along with a newly identified $RAC1^{A159V}$ hotspot mutation. This is a paralogous mutation to that in $K\text{-}RAS^{A146T}$ predominantly detected in colorectal carcinomas and shows the highest Rac1 activity in comparison to other existent Rac1 mutations[327].

1.2.2.2 Degradation defects

Abundance of Rac1 protein increases when its degradation is impaired. Rac1 is targeted for ubiquitin degradation by E3 ubiquitin ligases such as HACE1 (HECT-domain and Ankyrin-repeat Containing E3 ubiquitin protein ligase 1). This ubiquitin ligase catalyses its degradation once Rac1 is aberrantly activated by point mutations or by overactivation of upstream growth receptors like EGFR/ERBB2[328,329]. Such ubiquitin ligases are usually tumour suppressor genes and their downregulation lead to a functional neoplastic role.

1.2.2.3 Alterations in GEFs

Most of the Rac1 activities described previously require Rac1 to be GTP-bound. An accepted paradigm of Rho GTPases in cancer is that their level of activity prevails over their expression level. Guanine nucleotide exchange factor were initially discovered as oncogenes, given their transforming potential when microinjected in mammalian cells[330–332]. Two domains classify Rho GEFs into two categories, the Dbl RhoGEFs with DH and pleckstrin domains or the DOCK RhoGEFs. Some of the Rac1-specific RhoGEFs that are best characterised with cancer-related roles belong to the Dbl group and include Vav, Tiam1 or P-Rex[333]. In particular, these three RhoGEFs harbour binding regions on their sequence in order to be directly activated by extracellular stimuli prior their association with Rac1, like the activation of Ras or Src family tyrosine kinase to induce activation of Tiam1 and Vav, respectively[333].

Two isoforms for P-Rex exist, P-Rex1 and P-Rex2, both regulated upon PI3K/G-coupled receptor subunits Gβγ activation. These have been seen overexpressed in melanoma, breast and prostate cancer, promoting Rac1-dependent migration and invasion[334–336]. *P-Rex1*-deficient mice showed a metastatic-resistant phenotype in melanoma and prostate cancer and it mediates cancer progression in breast cancer by integrating ERBB2 receptor signals. Indeed, resistance to VEGFR inhibitors driven by P-Rex1-dependent activation of Rac1 was recently observed in prostate cancer stem cells[337].

Three different isoforms exist for Vav and all require a double phosphorylation step to be fully activated which is dependent on kinases like Src and Syk[333]. Src kinases are oncoproteins induced upon the activation of RTKs, linking its oncogenic role to substrates such as Vav. Vav proteins target all three studied RhoGTPases and the RTK-Src-Vav2-RhoGTPase axis of activation has been demonstrated to carry cancer-related activities in different contexts[258,338,339]. This axis promotes Rac1-dependent c-*myc* expression upon platelet-derived growth factor (PDFG) stimulation[339]. Likewise, when Vav is activated by other kinases such as Jak, it is able to drive invasion through matrix metalloproteinase 2 (MMP2) stimulation via the same axis[340]. Moreover, it was observed in head and neck squamous cell carcinoma (HNSCC) tumours that the high Rac1 activity responsible for their invasion was driven by EGFR-Vav2 activation[341].

Tiam1 is a Rac-selective GEF that can also indirectly activate Rac1 by functioning as a Ras effector[342]. The overexpression of *TIAM1* in a large number of cancer types including breast, colon and prostate strongly suggest its role as oncogene[343,344]. *Tiam1* deficiency in an Apc^{Min} colon cancer mouse model decreased tumorigenesis[345]. Indeed, it was identified as a WNT/β-catenin target gene given its increased expression upon *Apc* loss and TCF-binding sites upstream of the Tiam1 promoter. In a model for *Ras*-induced skin tumours, reduced *Tiam1* expression also decreased tumorigenesis and led to increased apoptosis, due to Ras-Tiam1 association and the lack of Rac1 activation by either Tiam1 or Ras[346]. Hence, Tiam1-Rac1 signalling seems relevant for both Ras- and WNT-driven tumours. Intriguingly, building evidence also suggests a negative role for Tiam1 in promoting invasion. For example, the few *RAS* and *TIAM1* double mutant skin tumours that grew in this study developed substantially more invasion than *TIAM1* wild type tumours[346]. The same scenario occurred in $Apc^{Min}TIAM^{-/-}$ mice, where invasion was enhanced compared to $Apc^{Min}TIAM^{+/+}$ control mice. In addition, Tiam1 expression was found to be significantly lower in advanced invasive breast tumours[345,347]. Recently, a mechanism was suggested whereby Tiam1 would function as a TAZ/YAP negative regulator. In the cytoplasm, Tiam1 binds to TAZ/YAP destruction complex and promotes their degradation. However, upon aberrant WNT signalling activity whereby the destruction complex is dissociated, Tiam1 translocates to the nucleus with TAZ/YAP and inhibits the initiation of their transcriptional activity involved in EMT, cell migration and invasion[348]. This highlights the complexity of RhoGEF functions and the need for better understanding of possible RhoGTP-independent roles yet to be described.

Regarding the DOCK type of RhoGEFs, fewer studies have been conducted compared to their fellow partners albeit it has greatly shown a similar and valuable outcome. Expression of *DOCK1*, one of the RhoGEFs group members, significantly correlates with poor prognosis in the most invasive subtypes of breast cancer. Besides, conditional deletion of *DOCK1* in a breast cancer mouse model protects mice from tumorigenesis and metastasis, indicating that DOCK1-Rac1 activation upon ERBB induction is necessary for tumour progression[349]. In a separate setting, Dock1 inhibition impairs both nutrient uptake and cellular invasion mediated by Rac1-dependent actin remodelling in the context of Ras mutation[350]. Further emphasising its important pro-tumorigenic role, a specific pharmacologically designed Dock1 inhibitor ablated the pro-metastatic functions of Dock1[350].

Collectively, the cancer-derived functions by aberrant RhoGEFs behaviour outlines their requirement for cancer transformation and suggest they are a key mechanism for Rac1 overactivation.

1.2.2.4 Alterations in GAPs

Following this logic, RhoGAPs, which enhance GTP hydrolysis and inactivate RhoGTPases, would be expected to act as tumour suppressors in cancer. Although downregulation of some has been demonstrated[351–353], recent unbiased studies have shown a pro-metastatic role for RhoGAPs such as CdGAP or RacGAP1[354–356]. This would suggest that, similar to Tiam1, functions other than their binding to and inactivation of RhoGTPases might account for these roles.

Therapeutic targeting of either RhoGEFs or RhoGAPs has been extensively investigated as an approach for some cancer treatments. Given that drug resistance has arisen for current RTK inhibitors mostly at the receptor level, tackling downstream cancer-hyperactivated pathways conjointly as an adjuvant may confer a potential benefit. Rac1 signalling is a promising anticancer target considering that 1) its activity is higher in neoplastic than in normal cells, 2) it modulates a broad range of biological process essential for cell transformation, 3) both upstream and downstream regulators of its activity appear to be druggable with promising positive outcomes. However, targeting such a ubiquitously expressed protein with many critical cellular outputs could also be detrimental to normal tissue.

1.2.2.5 Alternative splicing

The final mechanism for Rac1 overactivation is the production of a tumour-related alternative spliced isoform, Rac1b. Rac1b includes an additional exon in its sequence that causes a small but significant consequence to its conformation, leading to a virtually constitutive GTP-bound RhoGTPase. Since the alternative splicing process that leads to Rac1b formation seems enhanced in transformed cells, Rac1b may be a promising therapeutic target. The next section will focus on outlining its hitherto described functions, following a brief explanation of the RNA splicing process and other examples of malignant isoforms.

1.3 Alternative splicing

RNA splicing is a naturally occurring process whereby the exons that will make up the final mRNA product for protein translation are excised from the non-coding introns, therefore ensuring proper maturation of pre-mRNAs to mRNAs. The human body has a limited number of coding exons in the genome and alternative splicing (AS) acts by excluding or inserting additional exons, thus increasing proteome diversity. Over 130,000 alternative spliced isoforms can be generated from the 20,000 human protein-coding genes, with up to 95% of all multi exon genes alternative spliced[357]. According to Yang and colleagues, a sizeable proportion of alternative spliced isoforms share less than 50% of their protein interactome with their canonical isoforms, hence behaving virtually as distinct proteins and broadening protein functionality complexity[358].

The spliceosome is a macromolecular machinery composed of small ribonucleoproteins (snRNPs) and numerous auxiliary proteins. It mediates intron excision and exon joining by catalysing two sequential transesterification reactions. It recognises and binds to consensus sequences in the introns at the 5'splice site, 3'splice site (SS) and branch site (BS). It is targeted by RNA regulatory sequences termed exonic and intronic splicing enhancers (ESEs and ISEs), which are bound by SR protein-containing complexes that assemble to promote a splice site, or exonic and intronic silencers (ESSs and ISSs), that inhibit splicing via recruitment of heterogeneous nuclear ribonucleoproteins-containing complexes.

The spliceosome conducts both splicing and AS, and the decision to whether an exon is or is not spliced out is mediated, in a simplistic point of view, by the functional strength of the splice site and by the binding of regulatory proteins at enhancer and silencer sequences[359]. Two classes of transacting RNA binding proteins (RBPs) mediate positively or negatively spliceosome activity, heterogeneous nuclear RNPs (hnRNPs), which bind to silencers, or SR (Ser-Arg) proteins, which associate to enhancers. Their function is pivotal for facilitating splice recognition and their tissue-specific expression and context-dependent expression largely influence on leading to AS events and spliceosome activity. A cooperation of their functions also affects splicing in an antagonistic fashion, with the outcome (inclusion or exclusion) dependent on their concentration[360].

There are different types of AS, the most commonly occurring being cassette exon inclusion. Other kinds include alternative 5' or 3' splice site, intron retention, mutually

exclusive exon, exon scrambling or trans-splicing, the last two being rarely observed[361]. Of note, not all alternative spliced isoform products are functional and approximately one-third of premature mRNA are targeted for nonsense-mediated decay (NMD) degradation. Interestingly, this is also a mechanism AS uses to reduce translation of mRNA by promoting degradation of premature codons[362]. A dominant isoform transcript is produced per gene[363] but mutations in *cis*-acting splicing sequence, alterations in activity and protein level of regulatory proteins and mutations in the core splicing machinery might switch this dominant tendency towards less-frequent isoforms[364]. As the newly acquired functionality of certain gene might provide adaptation advantages in aberrant contexts, changes in their pattern of splicing is commonly associated with diseases, and cancer has been one of the best documented.

1.3.1 Implications in cancer

The large diversity that AS confers to the proteome can also be exploited by tumoral cells. For example, cellular processes altered in cancer cells, the so-called hallmarks of cancer, are all affected by aberrantly spliced isoforms (Figure 1.6)[365]. Intriguingly, some of the AS isoforms can have antagonist effects for example the apoptotic gene Bcl-x, whose splicing can produce two functionally opposite isoforms: Bcl-xS, a proapoptotic short form, or Bcl-xL, the antiapoptotic version[366]. The same scenario occurs for MCL1, another Bcl2 family protein that gives rise to a short pro-death isoform, MCL1S, and long antiapoptotic MCL1L form[367]. VEGFA is also a common example (Vascular endothelial growth factor A) which is normally secreted by endothelial cells to promote angiogenesis when oxygen is needed. Alternative spliced exon inclusion promotes the formation of a proangiogenic isoform over the antiangiogenic one, and a fine balance of their ratio dictates cells towards or against angiogenesis[368]. A non-antagonist example of AS is the tumour promoter gene CD44. This is a transmembrane cell adhesion molecule that translates into more than 20 distinct isoforms, two of them being commonly overexpressed in metastatic-related processes[369]. Therefore, mounting evidence exists of aberrant isoforms whose expression increases in transformed cells and serve as indicators of poor tumour prognosis.

A high frequency of hotspot mutations in splicing factor genes is also observed in some malignancies, such as in *U2AF1*, *SRSF2* or *SF3B1* genes, which translate into broad

transcriptome changes contributing to tumorigenesis[370]. Parallel screening shows that there are up to 30% more AS events in tumours compared to healthy tissue[371], indicating tumoral cells dependency on altered RNA splicing. These differences on AS isoforms suggest tumours have an intrinsically more activated spliceosome, which is also a consequence of a rapid protein translation. Therefore, this suggests that alternative splicing is an important tumorigenic process and reveals a vulnerability that may be differentially targeted from normal cells.

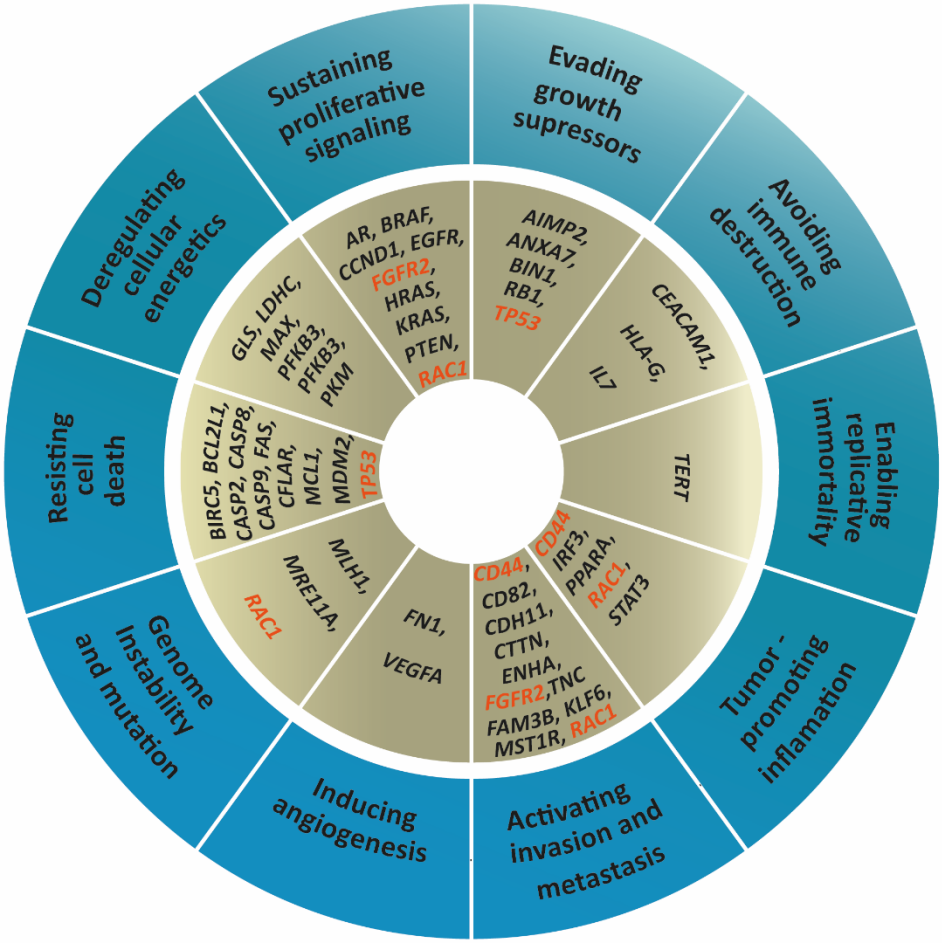


Figure 1.8: Schematic representation of genes whose aberrant splicing affects hallmarks of cancer processes. Orange genes highlight those involved in more than one hallmark. Adapted from *"Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes (Sveen et al, 2016)"*

1.3.2 Rac1b

Rac1b was discovered almost simultaneously by two different groups in 1999. Firstly, Jordan *et al.* noticed Rac1b for the first time when separating qRT-PCR products of *Rac1* cDNA extracted from a set of adult and foetal intestine, skin, liver and brain tissue[372]. On top of observing the expected 600bp band for Rac1, an additional band of 720bp was detected. Repeating the same analysis comparing colorectal tumours and normal mucosa showed that this band was preferentially expressed in tumour tissues, while Rac1 band intensity was fairly steady across both. Further sequencing revealed that the 720bp-band was the result of an additional insertion of 57 nucleotides between codons 75 and 76, which translated into 19 amino acids corresponding to a new undescribed exon, designated exon 3b (Figure 1.7A). Interestingly, the insertion occurs just after the switch II regions (residues 60-76), that controls the conformational change necessary for GTPase intrinsic activities. A few months later, Schnelzer *et al.* detected Rac1b in breast cancer and normal tissues by conducting qRT-PCR[373], confirming the discovery of a new alternative spliced isoform.

Rac1b is specifically yielded by Rac1 alternative splicing since neither Rac2 nor Rac3 have such exon 3b[374]. Despite being less expressed than Rac1, Rac1b presents higher activation. This activation was shown to be GEFs-dependent and it could also be negatively regulated by GAPs binding[375]. However, the additional 19 amino acids insertion prevents GDI binding, resulting in a steady constant activation. Further experiments concluded that Rac1b can be self-activated without GEFs due to a rapid nucleotide dissociation activity, and binding to the GEF Tiam1 or Vav2 do not enhance its activation[376,377]. Despite its binding to GAPs, it presents an impaired intrinsic GTP hydrolysis due to a highly mobile switch II, which favours its constant GTP-bound state (Figure 1.7B)[376].

When interrogating some of the main downstream features of activated Rac1, it was demonstrated that Rac1b cannot induce lamellipodia formation[373,375]. Moreover, it cannot activate the kinases PAK or JNK, but instead it can activate NFκB. Hence, it seemed like this additional exon insertion might confer a restricted and specific effector-binding site, which given its preferential tumoral expression, led to further investigation of its downstream signalling pathways.

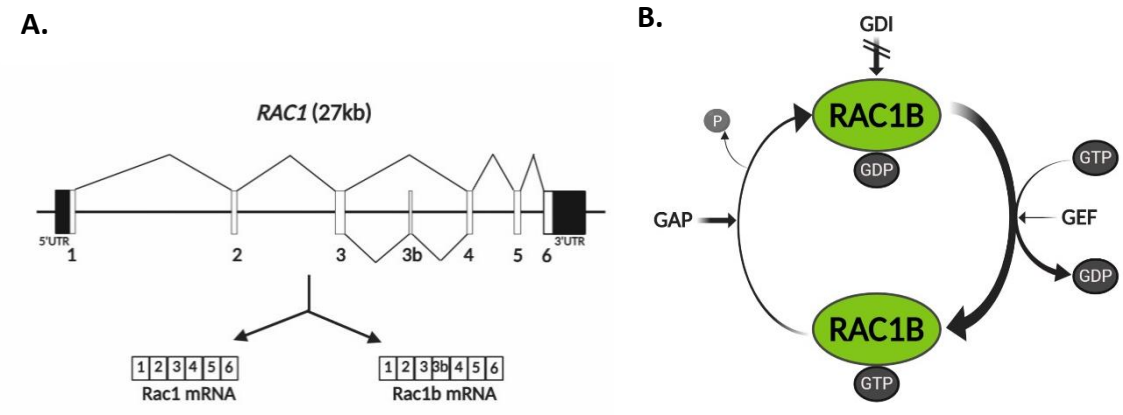


Figure 1.9: *RAC1B* transcript and protein activation overview. A: Schematic representation of the *RAC1* human gene and the two transcriptional products. B: Self-activating mechanism of Rac1b by its rapid intrinsic nucleotide exchange activity (although it can be enhanced by GEF binding), impaired GTPase reaction and inability to interact with GDI.
Figures adapted from Gonçalves *et al.* (2009) and Fiegen *et al.* (2003) respectively.

1.3.2.2 Cancer-associated functions

Given that Rac1 promotes cell adhesion and invasion, processes aberrantly activated during tumorigenesis, it was hypothesised that its constitutively active version would intensify its activity. From this regard, functional analysis of Rac1b has been assessed in comparison to Rac1 downstream effectors, although a partially different set of targets for Rac1b was already anticipated[373,375]. It was described that Rac1b interacts with $p120^{ctn}$, RACK1 and SmgGDS, which are proteins related with cell-cell adhesion, motility and transcriptional regulation[378], suggesting that Rac1b might be able to modulate cellular actin dynamics in a new route independent from PAK signalling.

Initial hints for Rac1b being involved in cell transformation was observed by Singh *et al.*, who noticed that Rac1b expression promoted loss of density-dependent inhibition of growth and acquisition of anchorage-independent growth in NIH 3T3 cells[377]. Moreover, Rac1b-expressing cells can grow under reduced growth factor and serum conditions compared to activated Rac1, probably because of its higher activation efficiency[379,380]. NF-κB activation is the main downstream pathway activated by Rac1b to promote cell growth. Mechanistically, Rac1b induces phosphorylation and membrane translocation of the NF-κB repressor IκBα, allowing NF-κB heterodimer to translocate to the nucleus[381]. In colorectal cancer cells, phosphorylation of IκBα is driven by Rac1b-dependent ROS production[382].

However, unlike Rac1 which can activate NF-κB by either releasing RelA (canonical pathway) or RelB (non-canonical), Rac1b can only promote the canonical route of NF-κB activation[383]. Rac1b-dependent activation of NF-κB is also relevant in papillary thyroid carcinoma, and both Rac1b expression and NF-κB activation associate with the presence of metastasis and poor survival[384,385]. G1/S cell cycle progression is stimulated by NF-κB-dependent and -independent expression of Cyclin D1. Rac1b can induce Cyclin D1 expression through interaction with Dishevelled-3 (Dvl3) and activation of the WNT signalling pathway[386]. Consistent with a tumour promoting role, it was shown that Rac1b can enhance already high levels of WNT signalling in tumours in a ligand-independent manner. According to their model, Rac1b recruits β-Catenin and Dvl3 and facilitates their translocation to the nucleus to interact with the promoter regions of c-Myc and Cyclin D1[387]. Rac1b, like Rac1, has a polybasic region (PBR) in its sequence that allows its translocation to the nucleus. In contrast to the suggested model for Rac1 as an essential component for nuclear translocation of β-Catenin, this model suggests that Rac1b is not required but it rather enhances its nuclear transport. This is the only evidence suggesting a nuclear role for Rac1b, as its constant GTP-bound status likely promotes a membrane localization. Further investigations would be interesting for uncovering its nuclear function. In addition, Li *et al.* have shown that Rac1b can promote c-Jun phosphorylation significantly more efficiently than Rac1, which then leads to cyclin D1 expression and cell cycle progression[380]. Phosphorylation occurs through Rac1b-dependent activation of JNK2. This contradicts previous observations, given that JNK and PAK are presumably the two Rac1 effectors Rac1b cannot activate[373,375]. However, they argue that Rac1b can only activate JNK2 but not JNK1 or JNK3, which could explain discrepancies with previous results. AKT2 has also been observed to be induced by Rac1b[377,380], which would stimulate anti-apoptotic cues by activating the anti-apoptotic protein MCL1.

Together with cell proliferation, migration and promotion of invasion are fairly well-accepted roles for Rac1b. EMT is commonly observed prior to cell invasion, albeit not being a strict requisite[388]. It consists of the activation of a transcriptional program (*Snail*, *Vimentin*, *Slug*, among others) for decreasing E-cadherin accumulation at the cell membrane. This allows morphological cellular changes for detachment from neighbouring cells and the surrounding extracellular matrix, hence enabling cell motility. In mouse mammary epithelial cells, it was observed that metalloproteinase 3 (MMP3) released from the tumour microenvironment, an essential protein for extracellular matrix proteins cleavage, activates Rac1 alternative splicing by releasing hnRNP A1 from exon 3b[389]. Rac1b protein then

promotes EMT by inducing the expression of Snail and Vimentin through ROS production. Simultaneously, ROS causes an indirect genomic instability due to DNA oxidative damage[390]. Rac1b seemed to be necessary for MMP3-derived activities, demonstrating the first evidence for Rac1b tumour invasion promotion. Other than in breast cancer, MMP3 activated Rac1b-driven invasion in human pancreatic carcinoma cells, and their expression was shown to correlate in PDAC human biopsies and associate with tumour grade[391]. A mechanism was suggested for MMP3-dependent EMT induction that depends on the cell substratum compliance. Initially, it was observed that cells undergoing EMT induced by MMP3/Rac1b axis needed to spread on a specific cell substrata, a requirement not observed for TGF-β-induced EMT[392]. Further analysis demonstrated that in collagen-rich substrata or stiff substrata, like those characteristic of tumours cells, fibronectin (an ECM protein) retains Rac1b in the plasma membrane to form a NADP complex and produce ROS for EMT induction. However, soft substrata prevents Rac1b membrane localisation by laminin-rich ECM, which sequesters Rac1b in the cytoplasm[393,394]. Rac1b IHC staining of breast tumour biopsies provided supporting evidence for this model, showing cytoplasmic staining of Rac1b in normal tissue while membrane staining was detected in tumour malignant biopsies.

In lung adenocarcinoma, EMT is observed in early adenomas in order to bypass oncogenic senescence and promote carcinoma formation. Interestingly, MMP3 induction of EMT through Rac1b expression was observed in lung adenocarcinoma cells whereby Rac1b would promote the expression of EMT markers such as *Vimentin* and decrease the expression of E-cadherin at the cell junctions. This will lead to EMT-bypass of senescence and consequently induce tumour progression[395]. This is an interesting observation of how the same stimulated process (EMT) could lead to two different phenotypes (senescence escape and cell invasion) but equally demonstrating the tumour progressive role of Rac1b.

Alternatively, Esufali et al. showed that Rac1b might downregulate E-cadherin expression through activation of its negative regulator Slug, leading to a decrease in cell adhesion[386]. They point to the PBR of Rac1b to mediate this function, since point mutations in this region abrogates Slug stimulation.

Recently, it was also suggested that Rac1b interacts with ArhGAP11A to promote invasion in hepatocellular carcinoma cells. Interestingly, this is a model for the aforementioned dual role of some RhoGAP proteins, which in certain circumstances might function as oncogenes. ArhGAP11A correlates with poor prognosis in hepatocellular

carcinoma biopsies. It interacts specifically with Rac1b, but not Rac1, for mediating its pro-tumorigenic functions which include enhanced cell growth, invasion and EMT[396].

Altogether, the available data in the literature consistently demonstrates a tumour promoting role for Rac1b in mediating invasion[379,390,395,391,386,394,392,397,377,384,396,398,399,400,390] (Figure 1.8A). However, a tumour suppressor role for Rac1b has also been shown by Ungefroren's lab. In apparent contradiction with Mehner *et al.*, they observed that non-malignant pancreatic ductal adenocarcinoma (PDAC) biopsies presented higher Rac1b staining compared to tumour samples. Functional analysis on PDAC patient-derived cells with *Rac1b* depletion showed an enhanced TGF-β-dependent migration, which diminished when *Rac1b* was overexpressed[401]. Negative regulation of TGF-β was mediated by depletion of SMAD3 phosphorylation and its transcriptional activity, which promotes cell migration and EMT upon activation. As it has also been suggested by other investigators, Rac1b and Rac1 might have antagonistic roles. However, while Nimnual *et al.* observed that Rac1b-driven Rac1 depletion led to an increase in Rho activity (due to their mutual negative regulation) and following promotion of invasion[397], Ungefroren *et al.* observed a decrease in invasion due to Rac1b inhibition of Rac1. In their model, a ratio of Rac1/Rac1b would dictate the invasive behaviour of a cell. This antagonistic relationship was also demonstrated in breast cancer cells, whereby Rac1 would mediate invasion and migration through increasing both the expression of *p21* and phosphorylation status of ERK1/2[402]. In addition to inhibiting the TGF-β /SMAD signalling cascade, Rac1b-mediated suppression of migration was also observed by inhibition of p38 and ERK1/2 MAPK phosphorylation[403]. Further mechanistic insights were recently revealed and suggest Par2 as the main Rac1b target[404]. Par2 is a G protein-coupled receptor that sustains activation of TβRI/ALK5. Initially, it was suggested that TGF-β-induced migration was driven by ALK5, as its expression increased upon *Rac1b* depletion and depletion of ALK5 partially attenuated the migratory phenotype[405]. However, Par2 overexpression can rescue Rac1b inhibition of TGF-β and its deletion phenocopies that of Rac1b. Moreover, increased ALK5 expression after *Rac1b* depletion was Par2-dependent, pointing to Par2 as the potential interacting protein with Rac1b (Figure 1.8B).

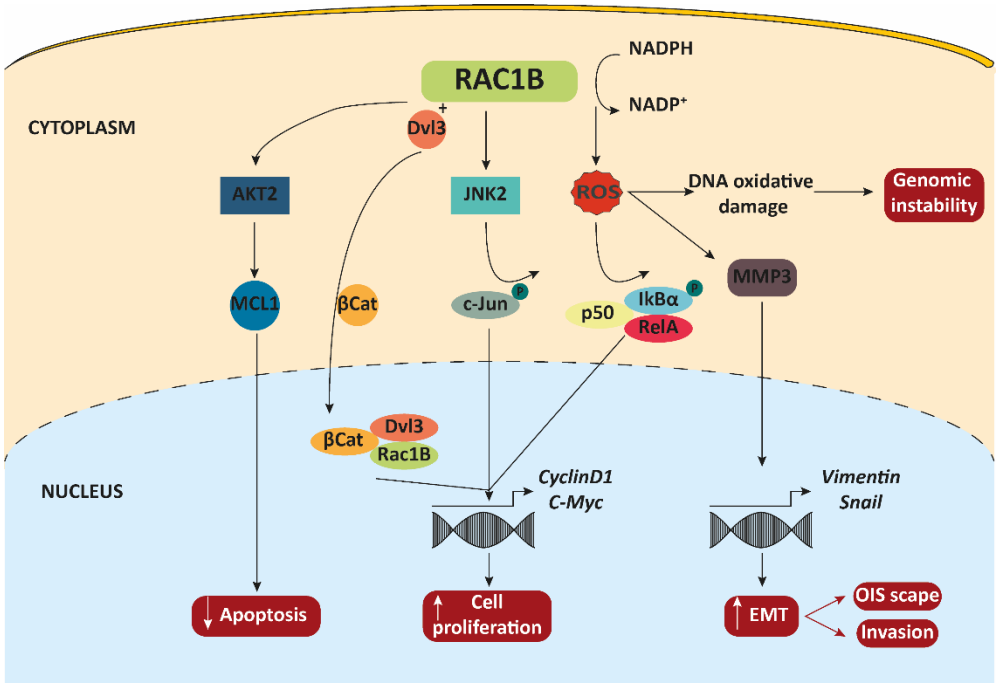
Rac1b expression has also been associated with other oncogenic processes. For example, it increases upon inflammatory episodes such as colitis and can promote COX2-dependent tumour growth in a colitis-associated cancer mice model[406]. Moreover, Rac1b can also promote the escape from oncogenic-induced senescence by suppressing the expression

of p14, p15 and p21 similar to the oncogenic escapes induced through EMT. Interestingly, Rac1b expression has largely been associated with CRCs driven by mutations in *BRAF*^{V600E} and a cooperation between Rac1b and *BRAF*^{V600E} was observed to sustain cell cycle stimulation and apoptosis[407].

Despite of the large efforts invested over the last 20 years to depict a comprehensive picture of Rac1b effectors and downstream activated signalling pathways, there is a clear lack of functional evidence *in vivo*. Li *et al.* deleted *Rac1b* by CRISPR in rat normal colon mucosa and showed an increase in apoptosis through AKT/MCL1 pathway[380]. However, in their previous *in vitro* experiments, Rac1b was able to phosphorylate c-Jun while no changes in JNK2 or c-Jun activation was observed in their *in vivo* model. This indicates that an *in vivo* translation and validation of the hitherto suggested roles for Rac1b is needed. To date, only three different studies have used *in vivo* mice models to modulate *Rac1b* expression in cancer. Stallings-Mann *et al.* overexpressed *Rac1b* specifically in murine lungs. These mice developed fibrosis which spontaneously led to tumorigenesis in a *Rac1b*-dependent dose[395]. Tumours were larger in size and more invasive than controls with wild type *Rac1b* and mechanisms include the EMT-bypassed senescence explained above. Zhou *et al.* used a *K-ras*-induced lung adenocarcinoma model to conditionally overexpressed *Rac1b* in the lung. They showed that, although Rac1b is not required for inducing *K-ras*-driven tumorigenesis, their co-expression enhanced tumorigenesis by increasing tumour number and size[408]. In contrast to the previous model, Rac1b could not promote tumorigenesis on its own, indicating that cooperation with other mutated cancer-related genes might be required to drive cell transformation. In the intestine, Kotelevets *et al.* showed that overexpression of *Rac1b* led to an increased proliferation through WNT signalling activation and an enhanced migration of intestinal epithelial cells in the small intestine[409]. It was insufficient to initiate tumorigenesis alone but overexpression of *Rac1b* in both the *Apc*$^{min/+}$ CRC model and in the *Il10*$^{-/-}$ colitis model led to an increased tumorigenesis in both models compared to controls.

These observations confirm the pro-tumorigenic role for Rac1b but do not clarify the underlying mechanism of actions that are driven by Rac1b. Given its described malignant roles in tumours and its biochemical nature that confers a constant Rac1-like activated signal, further investigations *in vivo* are of urgent need to unveil its potential benefit for therapeutic targeting.

A.



B.

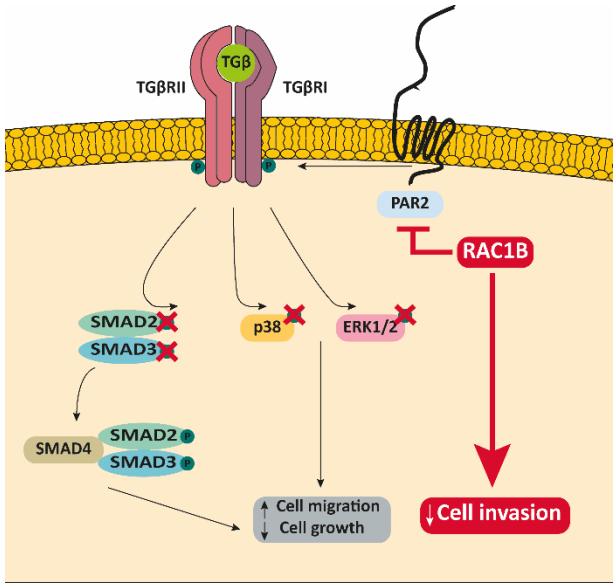


Figure 1.10: Downstream signalling pathways modulated by Rac1b and their cellular phenotypic translation. Figure A summarises the hitherto described functions of Rac1b related to pro-oncogenic processes, while figure B presents the suggested anti-oncogenic function of Rac1b driven by TGF-β downregulation.

1.3.2.1 Alternative splicing regulation

As discussed above, changes in the expression of splicing regulatory proteins can alter the splicing pattern of a gene. Hence, an accepted assumption for the shift from Rac1 to Rac1b production was the deregulation in the expression of Rac1 splicing factors, which are commonly overexpressed or downregulated in cancer[410,411]. Jordan's lab dissected for the first time the molecular factors responsible for Rac1b production and found that ASF/SF2 (SRSF1) splicing factor promotes inclusion of exon 3b, while SRp20 (SRSF3) skips it[412]. There are binding sites for both factors neighbouring exon 3b, acting as either enhancers (SRSF1) or silencers (SRSF3). Both factors function antagonistically and their ratio of expression leads to the dominant splicing event. Of note, signalling pathways upstream of these SR proteins indirectly modulate Rac1b splicing. They demonstrate that activation of the WNT/β-catenin signalling pathway, which SRSF3 is a target gene, downregulates Rac1b expression, whereas inhibition of the PI3K signalling pathway increases Rac1b expression through concomitant increase of SRSF1 at the protein and RNA level[412]. Protein kinases responsible for SR protein phosphorylation and activation include AKT2, AKT3, GSK3β and SRPK1, the latter being the key activator for SRSF1[413].

Another splicing factor which has been consistently demonstrated to modulate *Rac1* is hnRNP A1. Firstly, it was shown that hnRNP A1 binds to exon 3b and represses its inclusion. Indeed, an inverse correlation of hnRNP A1 and Rac1b was demonstrated by IHC analysis of breast tumour samples. It is suggested that the metalloproteinase MMP3 acts as an upstream regulator, releasing hnRNP A1-containing complex from the exon 3b and promoting its inclusion[389]. Another upstream signalling pathway affecting hnRNP A1 function is EGF/EGFR signalling. Mechanistically, EGF induces upregulation of SPSB1, an E3 ubiquitin ligase adaptor which targets hnRNP A1 for degradation and consequently promotes exon 3b inclusion and Rac1b expression, leading to EGF/EGFR-mediated cell migration[398]. Intriguingly, it has recently been demonstrated that a protein member from the hnRNP family can positively modulate Rac1b expression by favouring exon 3b inclusion. Xing and colleagues have shown that DIS3L2, an exoribonuclease with pro-tumorigenic roles in hepatocellular carcinoma cells, mediates hnRNP U recruitment to the exon 3b and both synergistically favour its inclusion. Moreover, Rac1b seems to be responsible for DIS3L2 oncogenic roles such as cell proliferation and tumour progression[399].

Finally, the last suggested splicing factor is ESRP1 or RBM35A. This, along with ESRP2, are considered tumour suppressors due to their TGF-β-dependent downregulation for promoting EMT. ESRP1 downregulates Rac1b expression through binding to a silencer sequence upstream the exon 3b[400]. In spite of being upregulated in tumours compared to normal tissue, ESRP1 is downregulated in the invasive front of the tumour, leading to Rac1b-driven invasion.

1.4 Animal models for CRC

Data from literature and previous section strongly suggest Rac1b is a functional promoter of tumorigenesis, but it is yet undescribed whether it functions primarily as a driver of tumour initiation or progression. The use of 2D tumoral cell lines to modulate the expression of a gene has the advantage of being a rapid and straight forward method to conduct assays, but its poor translatability to human tissues and difficulty to control background mutations limits its disease relevance. Instead, mice models have been the "gold standard" to mimic human cancer in the lab, which can also be the source of more sophisticated *in vitro* models like 3D organoids structures. Given the little references of *Rac1b* expression modulation in other systems than 2D cells[395,408,409], performing *Rac1b* gain and loss of function experiments in cancer mice models and 3D culture might considerably improve our current knowledge about its downstream targets.

Despite of the current debate on the type of clonal evolution initiates tumorigenesis, the Fearon and Vogelstein model of stepwise sequential accumulation of mutations is taken as reference when designing genetic strategies to emulate tumour stages in mice models. Therefore, mutation or loss of the *Apc* gene is usually taken as the key initial event for adenoma formation. Homozygous loss of *Apc* in the murine intestinal epithelial cells derives to a hyperplastic tissue with highly proliferative crypts[414], so heterozygous deletion is needed to model a long-term adenoma formation. In 1992 the first mouse model with *Apc* mutation was identified, the Apc^{Min} model (for Multiple Intestinal Neoplasms)[415,416]. It has a nonsense mutation at the codon 850 of the murine *Apc* gene, which is analogous to that of FAP patients. Homozygosity is embryonic lethal since the mutation is autosomal dominant, and the model is always heterozygous for *Min* ($Apc^{Min/+}$). When heterozygosity is lost over time, mice usually develop over a hundred tumours in the small intestine and rarely in the colon, with an average survival of 120 days[417]. Although being the most commonly model used for the last decades, drawbacks such as its specificity to the small intestine and the lack of invasive tumours due to a high and quick tumour burden led to the quest of more refined models. After different attempts of mutating the *Apc* gene[418–420] and other genes for either initiating tumorigenesis[421] or transforming adenomas into malignant tumours[422–424], a turning point in animal research was the introduction of genetic modified mice with spatiotemporal control[425,426]. Through tamoxifen-inducible Cre recombinase targeted to the intestinal epithelium by Villin[427] (small and large intestinal) or Carbonic anhydrase I (*Car1*)[428] (colon)

promoter, among other intestinal targetable promoters[419,429–432], researchers can decide the timing of the experiment and limit the gene modulation to a specific tissue.

Cre is a site-specific recombinase derived from the P1 bacteriophage that catalyses recombination between two *loxP* sites, its consensus DNA recognition sequence[425]. Since it was first discovered in 1984, there has been a huge effort to develop it as a genetic tool[433,426], and nowadays most mouse models commonly use tissue-specific Cre recombinase for genome alteration. Knock-out of a certain gene is achieved by inserting the *loxP* sites flanking a critical region of the gene, which will be "floxed" following Cre expression. In the same manner, gene overexpression can be achieved by inserting *loxP* sites flanking a stop cassette upstream of an exogenous sequence, allowing gene expression once recombinase is expressed. Soon after, geneticists improved this technology by creating an inducible recombinase with spatiotemporal expression control, such as the Cre-ER^{TM} fusion protein[434]. This mutant mouse estrogen receptor (ER^{TM}) fails to bind its physiological ligand (17β-estradiol) but it has high affinity for the synthetic 4-hydroxytamoxifen, the active tamoxifen metabolite. In normal conditions, ER^{TM} retains the fusion protein in the cytoplasm sequestered by Hsp90 and only after tamoxifen induction Cre-ER^{TM} is released and induces recombination in the nucleus[435]. In addition, when the Cre recombinase (either inducible or constitutively expressed) is combined with a tissue specific promoter, recombination can be targeted and delimited to a tissue[427].

In the present project, I will make use of the murine Villin promoter to target intestinal epithelial cells with a tamoxifen-inducible Cre (*VilCreER^{T2}*) to study *Rac1b* loss and gain of function in intestinal tissue homeostasis. Furthermore, I will combine *Rac1b* expression modulation along with *Apc* deletion to study its contribution in tumour formation and progression, and co-deletion with the tumour suppressor gene *TP53* to investigate its role in an invasive CRC model.

1.5 Concluding remarks

Colorectal cancer is the third leading cause of cancer-related death worldwide but it is also one of the most treatable cancers when diagnosed at early stages. Its survival has steadily increased over the last years, suggesting that improvements in CRC molecular characterisation has contributed to designing more efficient therapeutic treatments. However, a large proportion of patients present with invasive and metastatic disease with dismal prognosis and poor response to therapy. Therefore, there is an urgent need to identify new targets that could either serve as biomolecular indicators for anticipating pre-malignant tumours or as druggable candidates for therapy design.

The intestine is a highly proliferative tissue with a rapid cell turnover, required to avoid potential cellular damage due to its prolonged contact with abrasive luminal content. Signalling pathways that maintain intestinal homeostasis are finely tuned to restrict proliferation to the crypt bottom/stem cell zone. Cell differentiation is promoted above this region, whereby specialised intestinal epithelial cells undertake their metabolic and defensive functions. However, imbalance in this signalling network might lead to an expansion of the proliferation zone and the origin of hyperplastic intestinal foci or pre-adenomatous lesions. Key mutations that drive CRC tumorigenesis are mostly localised within these signalling pathways, which include the WNT, MAPK, TP53 and TGF-beta signalling pathways. In a neoplastic context, activating mutations in either the receptors or intracellular mediators of these pathways leads to a ligand-independent overactivation of signalling.

Connections between receptors and intracellular cues commonly require the association of small GTPases, which act as molecular switches and initiate signal activation upon ligand-receptor binding. Hence, overactivation of these switches causes a constant activation of their related pathways. The best example are the Ras small GTPases, whose mutated isoform K-Ras is observed in 40% of all CRCs. Rho GTPases, a subfamily group within the Ras superfamily, is also frequently altered in cancer. However, unlike Ras, mutations in their activating modulators rather than in the GTPase itself is more predominant. The activation of one of these Rho GTPases is achieved by the production of an alternative spliced isoform. This isoform of the Rac1 GTPase, namely Rac1b, possess an additional exon which causes conformational changes that prevents its interaction with the GTPase inhibitors GDIs

and enhances its intrinsic nucleotide exchange activity, conferring a virtually constant GTP-bound activation state. Interestingly, Rac1b confers a growth and invasive advantage over the canonical Rac1 protein isoform and tumours preferentially express higher levels of Rac1b compared to normal tissue. Despite extensive work on describing the pro-tumorigenic roles of Rac1b in colon, breast, pancreas, lung and thyroid cancer cells, there is a poor characterisation of its function *in vivo*. Given that Rac1b levels are very low in normal tissue, it stands as a potential candidate gene to target aberrant Rac1 activity observed in many cancer types. Therefore, I believe it is necessary to functionally characterise Rac1b *in vivo* and unveil, if any, the therapeutic benefit of its targeting.

1.6 Hypothesis and aims

Hypothesis

I hypothesise that Rac1b is a promoter of tumorigenesis and that its elevated expression in human CRC facilitates tumour progression and invasion.

Aims

In order to test the hypothesis, the aims of this project were as follow:

- Determine location of *Rac1b* expression in normal and tumoral mouse tissue.
- Investigate whether Rac1b is required for intestinal homeostasis.
- Investigate in human CRC samples whether *RAC1B* expression associates with poor clinical outcome.
- Evaluate whether *Rac1b* is essential for tumour progression by modulation of its expression in a benign adenoma mouse model.
- Determine whether *Rac1b* can promote or facilitate tumour invasion by modulating its expression in a more progressive CRC mouse model.
- Investigate the downstream effectors or pathways activated by Rac1b expression in the intestinal epithelium.
- If a pro-tumorigenic function is established, investigate the potential therapeutic benefit of targeting *Rac1b* expression.

Chapter 2: Material and methods

2.1 Mouse models

Mice were bred at the animal facilities of the University of Edinburgh and were kept in compliance with UK Home Office regulations. Mice were maintained in 12 hours light-dark cycles and had access to water and food *at libitum*.

All mice were crossed with mice carrying a tamoxifen-inducible Cre recombinase gene under control of the *Villin* promoter, to target intestinal cells[427]. All mouse lines were inbred within the same colony and had a mixed background.

2.1.1 Modulation of *Rac1b* expression

2.1.1.1 VillinCreER^{T2} Rosa26$^{lsl\text{-}Rac1b/lsl\text{-}Rac1b}$

VillinCreER^{T2} mice were crossed with mice carrying human *Rac1b* cDNA knocked in to the Rosa26-lox-Stop-lox locus[408]. Upon tamoxifen induction and Cre recombination, mice express the ectopic Rac1b cDNA. Recombination was evaluated by human *Rac1b* primers.

2.1.1.2 VillinCreERT2 Rac1b$^{fl/fl}$

LoxP sites were inserted on either side of exon 3b of the mouse *Rac1b* transcript (Rac1-202; ENSMUST00000100489.3), at positions 143,509,597 – 143,509,653. Upon Cre recombination, loxP sites were excised, preventing transcription of Rac1b protein. Mice were generated by Douglas Strathdee and David Stevenson (Transgenic Technology lab, Beatson Institute, Glasgow).

VillinCreER^{T2} Rosa26$^{lsl\text{-}Rac1b/lsl\text{-}Rac1b}$ (RAC1B^{OE}) and *VillinCre^{ERT2} Rac1b$^{fl/fl}$* (RAC1B^{KO}) mice were provided by Dr. Owen Sansom (Beatson Institute, Glasgow). For short-term assessment of *Rac1b* loss and overexpression in the murine intestine, mice were induced with two intraperitoneal (IP) injections of tamoxifen (120mg/kg at day 0 and 80mg/kg at day 1) and sacrificed 5 days after the first injection. For longer term evaluation of *Rac1b* overexpression,

RAC1B^{OE} mice were induced with two doses of tamoxifen IP and sacrificed 31 days post-induction. C57BL/6 wild type mice were used as controls. To assess tumorigenesis, both mouse lines, RAC1B^{OE} and RAC1B^{KO}, were crossed with the following *Apc^{fl}* and *Apc^{fl} TP53$^{fl/fl}$* alleles (Figure 2.1).

Tamoxifen solution (10mg/ml) was prepared by dissolving 1g of tamoxifen (Sigma, #T5648) into 10ml of pre-warmed 100% ethanol at 37ºC by discontinuous vortex. When dissolved, 90ml of corn oil (Sigma, #C8267) was added and solution was vortexed until dissolved. Stocks were kept at -20ºC.

2.1.2 Colorectal cancer mouse models

2.1.2.1 VillinCreER^{T2} Apc$^{fl/+}$

For specific loss of *Apc*, loxP sites were inserted into the introns around *Apc* exon 14, creating an *Apc* mutant allele (*Apc^{580S}*)[436]. Mice were provided by Dr. Owen Sansom and were crossed with RAC1B^{OE} and RAC1B^{KO} mice to create *VillinCreER^{T2} Apc$^{fl/+}$ Rosa26$^{lsl\text{-}Rac1b/lsl\text{-}Rac1b}$* (APC RAC1B^{OE}) or *VillinCreER^{T2} Apc$^{fl/+}$ Rac1b$^{fl/fl}$* (APC RAC1B^{KO}) mice. Mice were induced with a single dose of tamoxifen IP (80mg/kg) and were aged for tumorigenesis evaluation (section 2.1.1.4). *VillinCreER^{T2} Apc$^{fl/+}$* (APC) mice were used as controls. To assess the effect of acute *Rac1b* deletion in homozygous *Apc*-deficient cells, *VillinCreER^{T2} Apc$^{fl(fl}$ Rac1b$^{fl/fl}$* (APC$^{-/-}$ RAC1B^{KO}) mice were generated. These were induced with two doses of tamoxifen (120mg/kg at day 0 and 80mg/kg at day 1) and sacrificed 5 days following the first injection. *VillinCreER^{T2} Apc$^{fl(fl}$* (APC$^{-/-}$) mice were used as controls.

2.1.2.2 VillinCreER^{T2} Apc$^{fl/+}$ TP53$^{fl/fl}$

Loss of *TP53* was achieved by insertion of LoxP sites in introns 1 and 10 of the *TP53* gene. Upon Cre recombination, virtually all coding sequences (exons 2-10) were removed and p53 protein was not produced[437]. Mice were provided by Dr. Owen Sansom and were crossed with APC mice to create *VillinCreER^{T2} Apc$^{fl/+}$ TP53$^{fl/fl}$* (APC P53) mice. In turn, these were crossed with RAC1B^{OE} and RAC1B^{KO} mice and mice with the following genotypes were created: *VillinCreER^{T2} Apc$^{fl/+}$TP53$^{fl/fl}$ Rosa26$^{lsl\text{-}Rac1b/lsl\text{-}Rac1b}$* (APC P53 RAC1B^{OE}) and *VillinCreER^{T2}*

Apc$^{fl/+}$*TP53*$^{fl/fl}$ *Rac1b*$^{fl/fl}$ (APC P53 RAC1B^{KO}). Tumorigenesis assessment was performed by a single dose of tamoxifen (80mg/kg) in APC P53, APC P53 RAC1B^{OE} and APC P53 RAC1B^{KO} mice which were aged until endpoint signs were reached (details in section 2.1.2.3). For time-point evaluation of tumorigenesis, APC P53 and APC P53 RAC1B^{OE} mice were given a tamoxifen IP and sacrificed 31 days post-induction. Functional assessment of *Rac1b* overexpression on double *Apc*- and *TP53*-deficient cells was conducted on mice with the following genotype: *VillinCreER*^{T2} *Apc*$^{fl/fl}$ *TP53*$^{fl/fl}$ *Rosa26*$^{lsl\text{-}Rac1b/lsl\text{-}Rac1b}$ (APC$^{-/-}$ P53 RAC1B^{OE}) and *VillinCreER*^{T2} *Apc*$^{fl/fl}$ *TP53*$^{fl/fl}$ (APC$^{-/-}$ P53). These were induced with two tamoxifen IP injections (120mg/kg at day 0 and 80mg/kg at day 1) and terminated 4 days after induction.

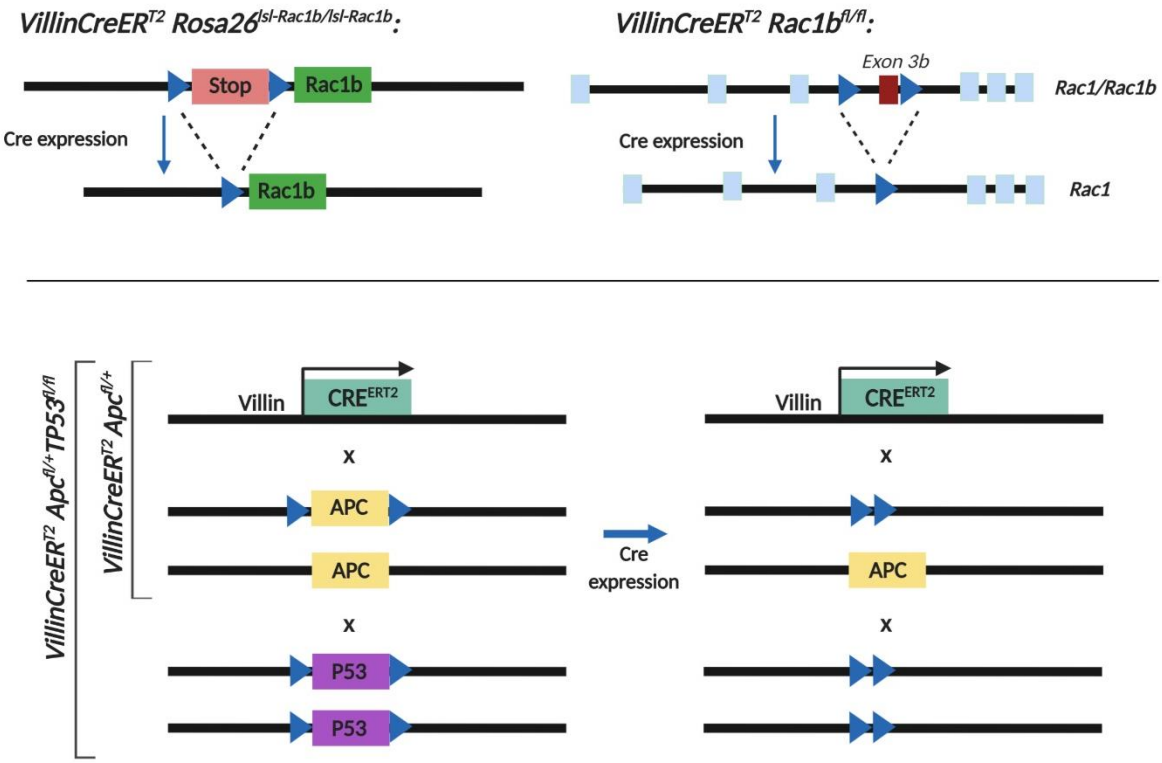


Figure 2.1: Mouse models for *Rac1b* expression modulation and CRC. Schematic genetic strategies for used genetic modified mouse models. Blue triangles represent loxP sites. Cre expression was achieved upon tamoxifen IP injection.

2.1.2.3 Disease surveillance and tumour quantification

After induction, mice were monitored twice a week until the end of the experiment. Disease progression was evaluated by the presence of rectal bleeding, weight loss, hunched position and pale feet, which is a sign of tumorigenesis-driven anaemia. Mice were terminated when these endpoints reached moderate severity. This was done in collaboration with Caroline Billard (ECRC, University of Edinburgh).

A BrdU IP of 200µl (Cell proliferation label, GE, #1392253) was given to the mice 2 hours prior to termination. Mice were sacrificed by cervical dislocation. Organs were harvested for evaluation of possible metastasis (spleen, pancreas, liver, lungs, lymph nodes and kidneys) and were fixed in 4% paraformaldehyde (PFA, VWR) for 24 hours at 4ºC. For tumour scoring, intestines were flushed in cold PBS, opened longitudinally on a piece of paper and submerged in methacarn fixative (60% Methanol, 30% Chloroform and 10% Glacial acetic acid). Number and size of the tumours was quantified. Tumour size was evaluated by measuring the surface area of tumours, and the sum of all area was considered the total mouse tumour burden. Alternatively, some mice within the cohort were fixed with PFA to allow following immunohistochemistry procedures.

2.2 Tissue histology

2.2.1 Fixation, tissue processing and sectioning

The intestine was harvested and flushed with cold PBS. For "Swiss-roll" sections, small intestine (SI) was cut in three pieces of approximately 10cm each. Both SI and large intestine (LI) were opened longitudinally and pinned on a wax plate. For PFA fixation, the tissue was covered in PFA in the plate for 24h at 4ºC. Afterwards, PFA was discarded, intestines were rolled and immersed in 70% ethanol. For methacarn fixation, tissues were immersed in methacarn for 15-30 minutes, rolled up, placed in formalin and kept at RT for 24 hours. For transversal parcel sections, 3 pieces of 1cm each of the SI were cut, stacked on tape and fixed for 24h at 4ºC in formalin. Tissues were switched to 70% ethanol prior to tissue processing

Tissue processing was automatedly conducted in the Tissue-TeK VIP Infiltration Processor (Sakura), which performs sequential incubation steps in ethanol (5 x 60min), xylene (3 x 60min) and paraffin (4 x 60min). Processed tissues were orientated and embedded in

paraffin wax and cooled down on a cold plate for several hours. Tissues were cut into sections of 5 µm thickness using the microtome (Leica), mounted onto glass slides (Thermo Fisher) and baked in an oven at 37ºC overnight (o/n). Prior to staining, sections were deparaffinised and rehydrated into serial washes of xylene and graduated ethanol dilutions until water (3 x 5min).

2.2.2 Tissue staining

2.2.2.1 Haematoxylin & Eosin (H&E)

Deparaffinised sections immersed in water were stained in Harris haematoxylin for 4:30 minutes and washed in running lukewarm tap water for another 5 minutes. Sections were differentiated in 1% acid alcohol for 5 to 7 seconds and quick washed in running tap water for 30 seconds. Sections were blued in saturated lithium carbonate for 7 seconds and washed for 5 minutes in running lukewarm tap water. Counterstain was conducted by immersing sections in eosin for 2:30 minutes, followed by 2-second wash in water to discard excess and 1min in 100% ethanol for dehydration. Sections were dehydrated in serial washes of ethanol (3 x 1:30min) and xylene (3 x 5min) and preserved with cover slides using DPX mounting medium (Sigma).

2.2.2.2 Periodic Acid Schiff (PAS)

Deparaffined sections hydrated in water were oxidized in 0.5% periodic acid solution (0.5g of Periodic acid, Sigma, dissolved on 100ml of distilled water) for 5min. Sections were rinsed in water, developed in Schiff reagent (Sigma) for 15min and washed for 5min in lukewarm running water. Counterstain was conducted by 1min immersion in Harris haematoxylin. Sections were washed, dehydrated as for H&E staining and mounted in DPX mounting medium (Sigma).

2.2.2.3 Immunohistochemistry (IHC)

Deparaffined and hydrated sections were placed in water before antigen retrieval (AR) process. Diluted AR solution (1:10, Citrate or EDTA buffer, Thermo Fisher) was heated up at 99ºC in a water bath. Sections were boiled into AR buffer for 20min (for citrate buffer) or 15min (for EDTA) and cooled down at RT for 30 min. Slides were washed and treated with 1.5% Hydrogen peroxide (Scientific laboratory supplies) diluted in water for 15min on a shaker. Following a water wash, sections were circled with a hydrophobic pen (StatMark pen, Sigma) and washed with PBST (1X Phosphate-buffered saline, IGMM Technical services, plus 0.1% Tween, Sigma). Slides were placed into a slide box and blocked with 5% goat serum (Sigma) at RT for 45minutes, depending on the antibody. Sections were covered with a small volume of diluted primary antibody (from 100 to 300µl per section) (Table 2.3) and incubated o/n at 4ºC. The day after, sections were washed with PBST (3 x 5min) and incubated with an HRP-polymer of conjugated mouse or rabbit secondary antibody (EnVision+ system, DAKO) for 1 hour at RT. Alternatively, signal was enhanced by incubating samples with Vectastain elite ABC-peroxidase kit (Vector laboratories), which involves 30min incubation at RT with a biotinylated secondary antibody followed by 30min incubation with an A+B (avidin plus biotinylated enzyme reagents) buffer solution. After 3 washes with PBST (5 minutes each), staining was developed by incubating sections with DAB substrate for 5min at RT. Sections were washed in PBST (3 x 5min), counterstained for 2min with haematoxylin, differentiated for 5 seconds with saturated lithium carbonate and washed for 5 min with running water. Before mounting with DPX mounting medium (Sigma), sections were dehydrated in ethanol and xylene.

2.2.2.3.1 Nuclear β-catenin IHC

Slight variations from the usual IHC protocol were conducted for nuclear and membrane staining of β-catenin. Hydrated slides were treated with a peroxidase block (4.16g Citric acid, 10.76g DiSodium Hydrogen Phosphatase 2 hydrate, 1g NaAz; add fresh H_2O_2 to 1.5%) for 45min on a shaker. Sections were then boiled for 50min in Tris EDTA buffer (242g Tris, 18.6g EDTA, diluted 1:5 in distilled water) and allowed to cool down for 1 hour at RT. After PBST washes (3 x 5min), sections were blocked with 1% BSA (Bovine serum albumin, Sigma) in PBS for 30min at RT and incubated for 2 hours at RT with the primary antibody

diluted in 1% BSA. Sections were then washed with PBST (3 x 5min) and incubated with an HRP-polymer of mouse conjugated secondary antibody (EnVision+ system, DAKO) for 1 hour at RT. Visualisation of positivity was carried out with DAB and slides were counterstained as previously explained.

2.2.2.4 BaseScope™

BaseScope™ probe was purchased from Advanced Cell Diagnostics and was designed for the transcript variant 1 of *Rac1* (NM_001347530.1) against the exon 3 – exon 3b junction (Figure 2.2), which includes the following nucleotides:

GTTGGAGACACATGTGGTAAAGATAGACCCTCCAGGGGCAAAGACAAGCCGATTGCCGACGTGTTC (underlined sequence refers to exon 3b).

Paraffin blocks were sent to Colin Nixon at the Beatson Institute for Cancer Research (Glasgow), who conducted BaseScope™ *in situ* hybridization. Analyses on stained slides were manually performed by scoring positive dots per crypt-axis unit (normal tissue) or intestinal area (hyperplasic tissue). Staining with a house keeping positive control probe was used as a reference for RNA integrity in tissues.

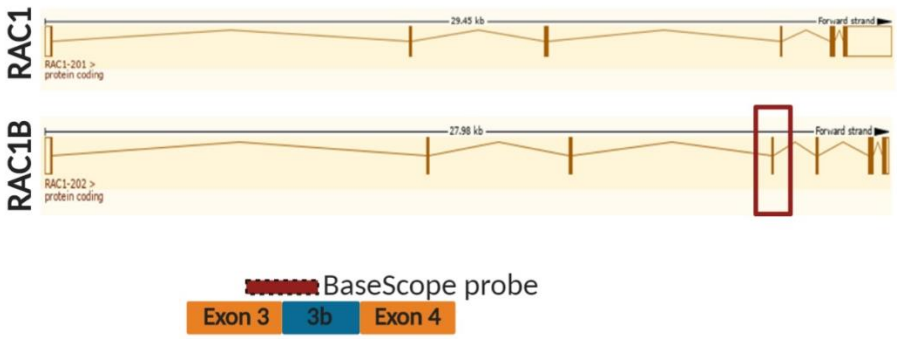


Figure 2.2: BaseScope probe for *Rac1b*. Schematic representation of BaseScope probe targeting exon 3 – exon 3b junction. Protein coding diagram downloaded from Ensembl.

2.2.2.5 QuPath IHC quantification

Slides were scanned using the NanoZoomer Digital slide scanner (Hamamatsu) and the viewer software NDP.view2 (Hamamatsu). QuPath software was downloaded (qupath.github.io) and scanned slides were imported to the software. The scoring area was delimited using the polygon tool, and cells within the area were scored by running "cell detection" option. For BrdU/proliferation scoring, positive cells were detected by running the "positive cell detection" option, selecting "nucleus: DAB OD max" option at the intensity threshold parameter and a threshold intensity of 0.5. Percentage of proliferation was calculated comparing positive cells to total cells detected within the area. For CC3/apoptosis scoring, positive cells were scored manually, and the percentage of apoptosis was calculated according to total cells detected in the area.

When using QuPath for histoscore analyses, positive cell detection was calculated for different threshold intensities. Score compartment would vary according to the staining pattern (nucleus, cytoplasm or cell). Based on the intensity of the staining, five different thresholds were chosen. The lowest threshold would include cells with the weakest staining intensity, while the highest would detect the most positive cells. Percentage of positive cells within each threshold would correspond to a score of 1 to 5. For histoscore analysis of SMAD4 IHC in tumours, threshold intensities were as follow: 0.12 (1), 0.24 (2), 0.36 (3), 0.48 (4) and 0.6 (5).

In contrast, area analysis (for tumour size, for example) was measured using the NDP.view2 software, and normal tissue BrdU/proliferation was scored manually.

2.3 Molecular biology techniques

2.3.1 RNA extraction

2.3.1.1 Whole tissue

When intestines were harvested, a piece of SI or LI was preserved in RNAlater (Sigma) and frozen at -80ºC. For extraction of RNA, a small piece of tissue (100-700µg) was dried from RNAlater solution and transferred to a new Eppendorf with RLT buffer (from Qiagen RNeasy Mini Kit) and 1% of β-mercaptoethanol (Sigma). Tissue was homogenised with stainless steel

beads (Qiagen) using the Qiagen Tissuelyser LT (Qiagen) at 50Hz for 5min. Lysate was centrifuged for 3min at 10,000*g*. Debris-free supernatant was transferred to a new Eppendorf and the same volume of 70% ethanol was added. Protocol was performed according to Qiagen RNeasy Mini Kit instructions. RNA was eluted in 50µl of RNase-free water from the kit and kept at -80ºC.

2.3.1.2 2D cells

Media was removed from wells and cells were washed twice with PBS. PBS was aspirated and 500µl of TRIzol (Invitrogen) was added per well (for 6-well plates). Cells were harvested by scraping the well with a P1000 tip and lysate was collected and left for 2min at RT to allow cell lysis. An equal volume of 100% ethanol was added and the homogenate was transferred to Zymo-Spin columns (Quick-RNA Microprep kit, Zymo Research). Protocol was followed as manufacturer's instructions indicated and RNA was eluted in ≥6µl of RNase-free water.

2.3.1.3 Intestinal organoids

Organoids were collected in a 15ml falcon tube by scraping with a P1000 tip and centrifuged at 300*g* for 3min. Supernatant was aspirated and 5ml of 1X PBS was added. Organoids were washed by pipetting up and down a few times. Cells were centrifuged again at 300*g* for 3min and this step was repeated twice. Matrigel-free pellet was resuspended in a lysis buffer (300µl of RTL, Qiagen, and 3µl of β-mercaptoethanol) and was left for 3min at RT to allow cell lysis. Homogenate was centrifuged at 10,000*g* for 3min and the protocol proceeded using the Qiagen RNeasy Mini Kit as in section 2.3.1.1.

Once RNA was eluted, residual DNA was removed using the DNA-free DNA removal kit (Ambion). Final RNA concentration and quality was evaluated by the Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific), using 2µl of RNA and 2µl of the eluted water as blank. An UV absorption ratio of A_{260}/A_{280} above 1.80 was considered optimal.

2.3.2 cDNA synthesis

Reverse transcription of RNA was conducted using the qScript cDNA SuperMix reagent (Quanta Bioscience). An amount of 1µg (for tissue) or 0.5µg (for cells) of RNA was combined with 4µg of the RT supermix, in a final volume of 20ul. PCR protocol is illustrated in Table 2.1 and was conducted using the Peltier Thermal Cycler (PTC-225, MJ Research Inc). PCR product was diluted 1:10 with RNase-free water and was kept at -20ºC for further use.

2.3.3 Real time PCR (qRT-PCR)

Amplification of cDNA was conducted using the CFX Connect Real-time System (Bio-Rad) machine. qRT-PCR primers were designed and purchased from Integrated DNA Technologies (IDT) (Table 2.2). A forward and reverse primer mix was diluted 1:10 in RNase-free water and reaction was performed using the LightCycler 480 SYBR Green I Master Mix (Life Technologies). Reactions were conducted in a 96-well plate (Axygen) and each reaction mix was composed of 10µl SYBR Master mix, 5µl cDNA, 1µl primer mix and 4µl RNase-free water. Reactions were conducted in duplicated as technical replicates. RNase-free water was used as a non-template control and β-actin was used as housekeeping gene in all reactions. qRT-PCR cycle conditions are illustrated in Table 2.1.

cDNA synthesis			
	Cycles	**Temperature**	**Time**
Incubation	1	25ºC	5min
DNA polymerization	1	42ºC	30min
Enzyme deactivation	1	85ºC	5min
RT-PCR			
Pre-incubation	1	95ºC	15min
Amplification	44	95ºC	10sec
		60ºC	30sec
		72ºC	30sec
Melting curve	1	65ºC - 95ºC	+0.5ºC/5sec

Table 2.1: PCR protocol for cDNA and qRT-PCR programs.

Gene	Forward primer (5'-3')	Reverse primer (5'-3')
Axin2	TGACTCTCCTTCCAGATCCCA	TGCCCACACTAGGCTGACA
BIM	GACAGAACCGCAAGGTAATCC	ACTTGTCACAACTCATGGGTG
Ccnd1	GCGTACCCTGACACCAATCTC	CTCCTCTTCGCACTTCTGCTC
Cited1	AACCTTGGAGTGAAGGATCGC	GTAGGAGAGCCTATTGGAGATGT
C-Myc	ATGCCCCTCAACGTGAACTTC	CGCAACATAGGATGGAGAGCA
Ctnnb1	ATGGAGCCGGACAGAAAAGC	CTTGCCACTCAGGGAAGGA
Defa20	TGTAGAAAAGGAGGCTGCAATAG	AGAACAAAAGTCGTCCTGAGC
E-Cadherin	CAGGTCTCCTCATGGCTTTGC	CTTCCGAAAAGAAGGCTGTCC
EphB2	GCGGCTACGACGAGAACAT	GGCTAAGTCAAAATCAGCCTCA
Etv4	CGGAGGATGAAAGGCGGATAC	TCTTGGAAGTGACTGAGGTCC
Etv5	TCAGTCTGATAACTTGGTGCTTC	GGCTTCCTATCGTAGGCACAA
Fzd3	ATGGCTGTGAGCTGGATTGTC	GGCACATCCTCAAGGTTATAGGT
hRac1b	CAGTAAGGGAGCTGCAGTGG	GCGAAGAGTTTGTCCTCAACC
IGFBP2	CAGACGCTACGCTGCTATCC	CCCTCAGAGTGGTCGTCATCA
Klk1	CAATGTGGGGGTATCCTGCTG	GGGTATTCATATTTGACGGGTGT
Krt20	AGTTTTCACCGAAGTCTGAGTTC	GTAGCTCATTACGGCTTTGGAG
Lef1	TGTTTATCCCATCACGGGTGG	CATGGAAGTGTCGCCTGACAG
Lgr5	GAGTCAACCCAAGCCTTAGTATCC	CATGGGACAAATGCAACTGAAG
Lyz1	GAGACCGAAGCACCGACTATG	CGGTTTTGACATTGTGTTCGC
mRac1b	TGTGGTAAAGATAGACCCTCC	CCCACGAGGATGATAGGAGT
Muc2	AGGGCTCGGAACTCCAGAAA	CCAGGGAATCGGTAGACATCG
Neu3	AGCAGGAAGAACAGAGTGGG	CCTCCATCAGTAGCCGTTGG
Olfm4	GCCACTTTCCAATTTCAC	GAGCCTCTTCTCATACAC
p15	TCTGCAGCTGGATCTGGTCC	TCCTGAAAGGTAGAGGGCCC
p21	CCTGGTGATGTCCGACCTG	CCATGAGCGCATCGCAATC
Rac1	GAGACGGAGCTGTTGGTAAAA	ATAGGCCCAGATTCACTGGTT
Slc14a1	TTAAAGTAGACCGGGGTGAAAAC	ACCCGTGACGTAGCCAAGTA
SMAD4	ACACCAACAAGTAACGATGCC	GCAAAGGTTTCACTTTCCCCA
SMAD7	GGCCGGATCTCAGGCATTC	TTGGGTATCTGGAGTAAGGAGG
Tcf7	AGCTTTCTCCACTCTACGAACA	AATCCAGAGAGATCGGGGGTC
TGF-β1	CTCCCGTGGCTTCTAGTGC	GCCTTAGTTTGGACAGGATCTG
TGF-β2	CTTCGACGTGACAGACGCT	GCAGGGGCAGTGTAAACTTATT
Tiam1	GAAGCACACTTCACGCTCC	CTCCAGGCCATTTTCAGCCA
Tnfrsf19	TTCTGTGGGGGACACGATG	AGAAAATTCAGCGCAGATGGAA
Trpv6	AGGGGTTAATACTCTGCCTATGG	GCACCTCACATCCTTCAAACTT
TβRII	CCGCTGCATATCGTCCTGTG	AGTGGATGGATGGTCCTATTACA
Znrf3	GGCGACTATACCACCCACAC	CTTCACCACTCCTACCCAGC
β-actin	GTGACGTTGACATCCGTAAAGA	GCCGGACTCATCGTACTCC

Table 2.2: **Forward and reverse primer sequences used for qRT-PCR.**

2.3.4 RNA sequencing

RNA quality was assessed by Stewart McKay (HGU, University of Edinburgh) on an Agilent 2200. Samples were considered optimal when RIN (RNA integrity number) values were above 8.

An amount of 1µg of RNA was sent to GATC Biotech AG (Germany) for library preparation (random primary cDNA library) and Illumina RNA sequencing (a single read of 50bp length and 30 million reads guaranteed per sample (+/-3%)). GRCm38 (Ensemble 95) was utilised to define coding and non-coding cDNA sequences and to build a transcriptome index for Kallisto, which was used to quantify the transcript abundances. Estimated counts were imported with Tximport and differentially expressed genes were identified using both Cuffdiff and DESeq. Sequencing data was analysed by Dr. Stuart Aitken (IGMM, University of Edinburgh). Genes with a p-value higher than 0.05 were considered for pathway analyses using Innate DB database (innatedb.com).

2.3.5 Protein extraction and quantification

2.3.5.2 2D cells

From cells cultured in a 6-well plate, media was removed and wells were washed twice with PBS. A volume of 100µl of RIPA buffer (Sigma) supplemented with 1µl of Phosphatase inhibitor cocktail 3 (Sigma) and 1µl of Protease inhibitor cocktail (Sigma) was added at each well. Cells were scraped with a P1000 tip, collected on an Eppendorf and left on ice for 10min to aid cell lysis. Proteins were clarified by a 30' vortex and a centrifugation step at 10,000*g* for 10min at 4ºC.

2.3.5.3 Intestinal organoids

Organoids in culture were collected and washed, and Matrigel was removed by sequential PBS washes and centrifugation steps. However, the pellet was resuspended with a volume of RIPA lysis buffer according to the size of the cellular pellet (from 100µl to 500µl). Lysate was incubated on ice for 10min and procedure followed as for 2D cells protein extraction.

Protein quantification was conducted using the Pierce BCA protein assay kit (Fisher). Following manufacturer's protocol, 200µl of working solution was incubated with 10µl of protein in a 96-well plate with a flat bottom (Greiner) and each sample was conducted in triplicate. After 30min of incubation at 37ºC, absorbance of the plate at 562 nm was read and protein concentration was calculated according to the albumin standard curve.

2.3.6 Western blotting

Whole protein extracts were evaluated by Western blot. Protein lysates were separated on 4-12% Bis-Tris protein gels from NuPage (Thermo Fisher). A mix of protein lysate (6µg-10µg), 4X of sample loading buffer (Thermo Fisher) and 10X of DTT (Sample reducing agent NuPage, Thermo Fisher) was prepared for each sample and was denatured at 99ºC per 5min. The gel was placed in the tank filled with 1X MOPS (for large proteins) or MES (for small proteins) running buffer (Thermo Fisher) and proteins were loaded into the wells. Running was carried out at 120V at RT and, when finished, the gel was transferred to a methanol pre-activated nitrocellulose membrane (Fisher Scientific). Diluted 20X Transfer buffer (Thermo Fisher) in water and methanol was used and transfer was conducted at 4ºC for 1:30h at 250mAmp. Transfer efficiency was evaluated with Ponceau solution (Sigma) and membrane was washed with PBST 3 times prior to blocking with 5% milk/PBST (dried milk, Marvel) for 1 hour at RT. Alternatively, for probing phosphorylated proteins, 3% BSA/PBST was used for blocking. Blocked membrane was incubated with 1º antibody diluted in blocking solution at 4ºC o/n (Table 2.3 for antibody concentrations). After incubation, membrane was washed with PBST (3 x 5min) and was incubated with 2º antibody diluted in blocking solution for 1 hour at RT (anti-rabbit or anti-mouse IgG, HRP-linked, Cell Signalling). Following 3 more washes with PBST, signal was detected by chemiluminescence using ECL Plus substrate (Thermo Scientific) for 5 min and membranes were develop with ECL hyperfilm (Amersham) at the darkroom using the Medical Film Processor SRX-101A (Konica Minolta).

To strip and for further probing with other antibodies, membrane was incubated for 15min at RT with a Stripping solution (1:10 dilution, Millipore).

Densitometry of Western blot bands was assessed using ImageJ (https://imagej.nih.gov/ij/download.html).

Antibody		IHC			WB		
Name	Cat#, Company	Concentration	AR	2nd Antibody	Concentration	Block	2nd Antibody
AKT	#9272, Cell Signalling				1:2000	5% milk	Rabbit HRP
BrdU	#bs-0489H, Bioss	1:500	Citrate	Mouse HRP			
C-Caspase 3	#AF853, R&D	1:800	Citrate	Rabbit ABC			
CD3	#A045229, DAKO	1:500	EDTA	Rabbit HRP			
E-Cadherin	#3195, Cell Signalling	1:200	Citrate	Rabbit HRP			
EGFR	#2232S, Cell Signalling				1:1000	5% milk	Rabbit HRP
ERK1/2	#4695, Cell Signalling				1:2000	5% milk	Rabbit HRP
F4/80	#132102, Biolegend	1:100	Proteinase K	Mouse ABC			
Lysozyme	#A009, DAKO	1:1000 (1h RT)	Citrate	Rabbit ABC	1:3000	5% milk	Rabbit HRP
Mucin 2	#GTX100664, Genetex	1:750	Citrate	Rabbit HRP			
Myc-tag	#2276S, Cell Signalling				1:10,000	5% milk	Mouse HRP
p-AKT	#4060S, Cell Signalling				1:3000	3% BSA	Rabbit HRP
p-EGFR	#2234S, Cell Signalling				1:500	3% BSA	Rabbit HRP
p-ERK1/2	#4370S, Cell Signalling				1:3000	3% BSA	Rabbit HRP
pH2AX	#9718P, Cell Signalling	1:100	Citrate	Mouse HRP			
pSMAD3	#ab52903, Abcam				1:5000	3% BSA	Rabbit HRP
Rac1b	#09-271, Millipore	1:2000 (1h RT)	EDTA	Rabbit HRP	1:1000	5% milk	Rabbit HRP
SMAD4	#sc-7966, Santa Cruz				1:2000	5% milk	Mouse HRP
Streptavidin (HRP)	#ab7403, Abcam				1:10,000	3% BSA	
Vinculin	ab73412, Abcam				1:5000	5% milk	Rabbit HRP
β-actin	#4970, Cell Signalling				1:5000	5% milk	Rabbit HRP
β-catenin	#610154, BD	1:50 (2h RT)	Custom	Mouse HRP	1:5000	5% milk	Mouse HRP
BIM	#2933, Cell Signalling				1:1000	5% milk	Rabbit HRP

Table 2.3: List of antibodies used for IHC and WB. Incubation time were o/n unless stated. AR Proteinase K: 20X Proteinase K (Sigma) dissolved 1:20 in TE buffer, incubated for 5min at 37ºC.

2.3.7 Protein interaction study by BioID

2.3.7.1 Cloning the BirA enzyme into a lentiviral vector*

Myc-BioID2-MCS plasmid was purchased from Addgene (#74223), containing the BirA with a myc tag on the C-terminal. Bacterial stab was purified by Miniprep (Qiagen) following manufacturer's instructions. Myc-BioID2 sequence was introduced into the pLJM1-EGFP (#19319, Addgene) lentiviral vector by digesting 5µg of both plasmids with NheI-HF (NEB, #R3131) and EcoRI-HF (NEB, #R3101S) restriction enzymes (RE). CutSmart (NEB) was used as the digestion buffer and digestion protocol followed manufacturer's instructions. Digestion was separated on 1% agarose gel (3g agarose dissolved in 300ml 1X TAE buffer) ran at 120V at RT.

Bands appeared at 775bp for Myc-BioID2 and 7315bp for pLJM1 vector. These were cut out of the gel using the Zymoclean Gel DNA recovery kit (Zymo Research) and DNA was quantified by Nanodrop. Ligation was performed using the T4 ligase kit (Promega) o/n. Ligation was transformed into One Shot TOP10 bacteria (Thermo Fisher) by heat shock,

followed by 1-hour incubation at 37ºC with SOC media (provided with bacteria) and plated in Ampicillin (AMP, IGMM Technical services) resistant plates o/n at 37ºC. Efficiently transformed colonies were picked and grew on 2ml L-Broth media (IGMM Technical services) supplemented with Penicillin-Streptomycin (PEN/STREP, IGMM Technical services) and AMP for following Miniprep.

2.3.7.2 Amplification of Rac1 and Rac1b by PCR

cDNA synthesised from normal murine SI was used to amplify *Rac1* DNA. On the contrary, *Rac1b* DNA was purchased from GeneScript (#OHu22224). Primers were designed to insert EcoRI sequence at the C-terminal and SacII sequence (NEB, #R0175S) at the N-terminal. Primer sequences and PCR protocol were as follows:

- RAC1/RAC1B F (5'-3'): TTTTTGAATTCCAGGCCATCAAGTGTGTG
- RAC1/RAC1B R (5'-3'): TATATCCGCGGTTACAACAGCAGGCATTTTCTC

	Cycles	Temperature	Time		For 200µl reaction
Pre-incubation	1	98ºC	4min	5X HF Phusion Taq buffer	40µl
Denaturation	27	95ºC	45sec	dNTP mix	4µl
Annealing	27	59ºC	45sec	10µM F+R primers	20µl
Elongation	27	72ºC	40sec	Phusion Taq	2µl
Elongation	1	72ºC	7min	Template (125ng)	2µl
Cooling	1	4ºC	end	Water	132.4µl

Table 2.4: PCR protocol for *Rac1* and *Rac1b* amplification.

2.3.7.3 Cloning of Rac1/Rac1b into the pLJM1-myc-BioID2 vector

Rac1 and *Rac1b* sequences were independently inserted into the lentiviral vector by using the pGEM-T Easy Vector Systems (Promega). This system required of an A-tailing reaction prior to ligation. Reaction was conducted as follows:

Temperature	Time		
		DNA	3µl
95ºC	1min	**Taq buffer**	0.5µl
70ºC	30min	**1mM dATP**	1µl
4ºC	end	**FastTaq**	0.5µl

Table 2.5: A-tailing protocol.

The reaction was ligated into the pGEM-T vector following manufacturer's instructions. Ligation was transformed as detailed in section 2.3.7.1 and transformation was plated in X-gal/IPTG plates (reagents and protocol from Goldbio, Blue White Screening using X-Gal and IPTG), necessary to discriminate between positive (white) and negative (blue) transformed colonies. White colonies were grown for Miniprep extraction and insertion was verified by digesting plasmid with EcoRI and SacII RE and running product into a 1% agarose gel. Detected bands were extracted using the Zymoclean Gel DNA Recovery kit and DNA was sequenced by Jeffrey Joseph (Technical Services, MRC, University of Edinburgh).

pLJM1-myc-BioID2 vector was digested with EcoRI and SacII RE and was purified after running on an agarose gel. Digested vector was ligated with pGEM-T inserts whose sequences were verified. Ligations were transformed into TOP10 bacteria and colonies grown, then purified by Miniprep. These were sent for sequencing and plasmids with verified sequences were amplified by Maxiprep (Qiagen). A vector with no insert was amplified by Maxiprep to use as an experimental control (Figure 2.3).

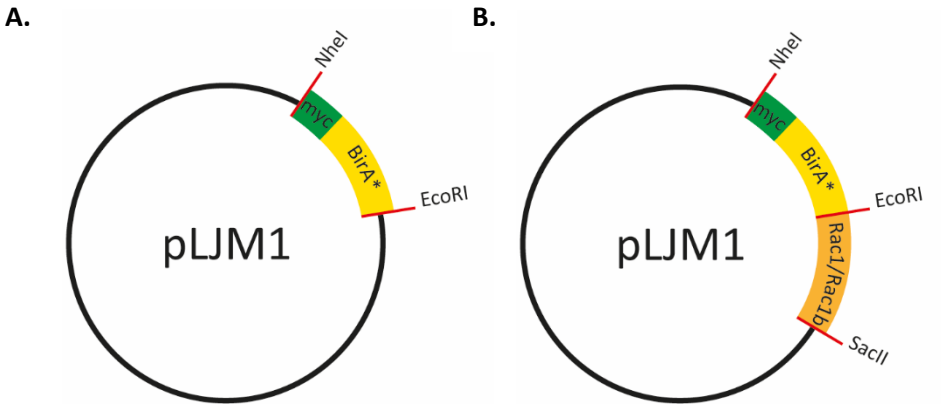


Figure 2.3: BioID plasmids. Schematic representation of the resultant cloned plasmids. A: Control for BioID experiment, with only the myc-BirA* sequence cloned. B: Either Rac1 or Rac1b were cloned into the pLJM1 lentiviral vector, N-terminal to BirA* enzyme.

2.3.7.4 Streptavidin-biotin pull-down for MS

Upon biotin treatment, transfected cells were collected for protein extraction. A volume of 500µl of RIPA buffer supplemented with phosphatase and protease inhibitors was added per 10cm^2 dish. Cells were scraped and transferred to a 1.5ml Eppendorf tube. Following 15min of lysis incubation on ice, cells were vortexed for 30'' and clarified at maximum speed for 10min at 4ºC. In the meantime, 15µl of Streptavidin Sepharose matrix (GE Healthcare, #17-5113-01) were neutralised on 1ml of RIPA buffer by gently pipetting, and centrifuged for 2min at 1,000*g*. RIPA supernatant was removed and clarified proteins were added to the beads. These were mixed without pipetting and incubated on a rotator at 4ºC for 6 hours.

After incubation, the mix was spun down at 1,000*g* for 5min. Without altering the beads pellet, supernatant was removed and 1ml of wash buffer (50mM TrisCl, 8M Urea, pH 7.4) was gently added and mixed by inversion. The tube was placed on a rotator for 8min at RT and centrifuged for 2min at RT. This step was repeated three times, the last wash being with urea-free wash buffer. Beads were kept at -20ºC without supernatant to send for MS analysis. MS experiment and analysis of the data was performed by Dr Alexander von Kriegsheim (Mass Spectrometry Facility Manager, IGMM, University of Edinburgh). Interactome analyses from the resultant protein hits was conducted using the NetworkAnalyst web-based platform (https://www.networkanalyst.ca).

For validation of BioID hits, pull-down beads were resuspended with 55µl of RIPA buffer supplemented with 5.5µl of 10X DTT. Lysate input was used as a positive control. Protein lysates were run on a Western blot as detailed in section 2.3.6.

2.4 2D cell culture

2.4.1 Cell line model

The mouse rectal carcinoma cell line CMT93 was used for 2D cell culture experiments. Cell line was previously characterised by Kitamura *et al*[438] (Table 2.6).

	Ccl9	Wnt	TGFβ	K-Ras	PIK3/AKT	p53	B-Raf	Other
CMT93	+++	Intact	Intact	Not mutated	Activated	Not mutated	Not mutated	STAT3 activated

Table 2.6: Mutational profile CMT93 cell line. Adapted from Kitamura *et al.*

2.4.2 Routine cell culture maintenance

Cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) (Sigma) supplemented with 10% foetal calf serum (FCS, IGMM Technical services) and 1% penicillin and streptomycin (Life technologies) and cultured in T25 or T75 ($25cm^2$ or $75cm^2$) flasks (Corning) at 37ºC in a 5% CO_2 incubator (Heracell 150i, Thermo Fisher). Manipulation of cells was carried out under a laminar flow cabinet. To passage, cells were washed with PBS and a volume of 0.25% Trypsin-EDTA (Gibco) was added to detached cells from flask. When all were detached, DMEM was added to neutralise Trypsin. Cells were recollected into a falcon tube and were centrifuged for 5min at 300*g*. Cells were diluted 1:5 to 1:10, depending on the experimental plan.

To seed a precise number of cells, Countess™ cell counter was used (Thermo Fisher). Following the centrifuge step, pellet was resuspended with 5ml of DMEM. In a new Eppendorf, 10µl of trypan blue (Thermo Fisher) was merged with 10µl of cells and the mix was inserted into a disposable Countess chamber slide. The slide was introduced into the instrument and counting was automatedly conducted.

To maintain cell stock, cells were frozen at a concentration of 3.5-5milions cells/ml diluted in 10% DMSO (Sigma). Aliquots were frozen in an isopropanol freezing chamber (MrFrosty, Thermo Fisher) at -80ºC prior transfer to N_2. Cells were recovered by thawing for 1-2min in a 37ºC water bath.

2.4.3 Plasmid transfection

One day before transfection, cells were seeded in a 6-well plate at a density of 3×10^5 cells/well. 4µg of DNA plasmid per well was diluted in 250µl of Opti-MEM (Gibco). In a separate tube, 25µl of Lipofectamine 2000 was gently mixed with 150µl of Opti-MEM, and left at RT for 5min. The DNA-containing tube and Lipofectamine-containing tube were mix and incubated at RT for 20min. Cell media was replaced with 2ml fresh DMEM and 500µl of transfection mix media. Cells were incubated at 37º and transfection was evaluated 48 hours later.

2.4.4 Biotin treatment for MS analysis

To start, a total of ~3 x 10^6 cells were seeded on two 10cm^2 dish (Thermo Fisher). The following day, transfection was carried out as detailed in section 2.4.3, using 24µg of DNA and 60µg of Lipofectamine 2000 per dish. 40 hours after transfection, transfection media was replaced by 10ml with 50µM of Biotin (Sigma; from 1mM stock, 0.0122g Biotin dissolved in 50ml DMEM). Cells were collected 20 hours later for protein extraction (section 2.3.5.2).

2.4.5 Depletion of *Rac1b* with morpholino

An antisense morpholino to target exon 3b exon skipping was especially designed by Dr Linda Popplewell and Prof George Dickson (Royal Holloway, London) and was purchased from GeneTools (USA).

To deplete *Rac1b* cells were seeded for 2 days with 5µl/ml of *Rac1b* Morpholino (*Rac1b* PMO) or non-targeted morpholino (NT, control) in the media (Figure 2.4). After two days, cells were split and counted with an automated cell counter. In a 6-well plate, 2.5 x 10^5 cells were seeded per well with 2.5µl of *Rac1b* PMO or NT for one day. Depletion was determined 24 hours later and cells were ready for further experiments.

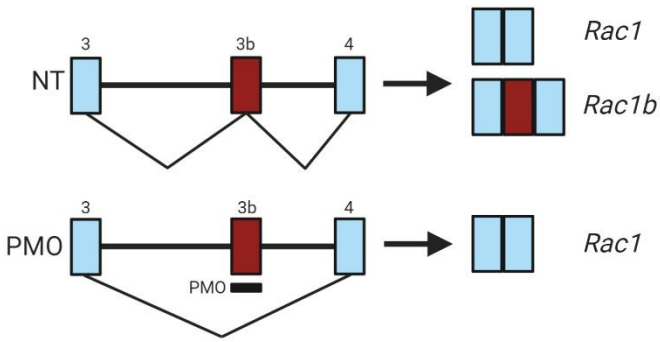


Figure 2.4: Schematic representation of morpholino-mediated *Rac1b* knockdown. The morpholino (PMO) promotes exon skipping by binding to the exonic splice enhancer sequence in exon 3b.

2.4.6 EGF stimulation

On the day prior to experiment, 2.5 x 10^5 cells/well were seeded in a 6-well plate. The day after, cells were serum starved for 4 hours before EGF was added to the media (except the non-starved control cells). A media mix solution was prepared with 20ng/ml of EGF (Invitrogen). After starvation, cell media was replaced by 2ml of EGF-containing media. Plates were put back into the incubator and cells were collected for protein extraction after 5, 10, 15 and 30min. Two EGF non-treated wells were used for treatment control and to validate *Rac1b* depletion by qRT-PCR.

2.5 3D Organoid culture

2.5.1 Small intestine epithelial isolation and culture

Small intestine was harvested and flushed thoroughly with cold PBS. The tissue was opened longitudinally and villi were scraped off with a coverslip. Under the hood, small pieces of tissue (~5mm) were cut and transferred to a 50ml falcon tube. Tissue was washed with 10ml of cold PBS up to 5 times, allowing the pieces to settle down and discarding the supernatant with debris. A volume of 25ml of 2mM EDTA (IGMM Technical services) in PBS was added to the intestinal pieces, and the falcon was incubated at 4ºC with agitation. This step allows the crypts to separate from surrounding tissue. After 30min, EDTA was aspirated and pieces were gently washed from the EDTA with 10ml of PBS. Supernatant was discarded and 10ml of cold PBS was added to the pieces. Tissue was thoroughly pipetted up and down about 10 times to loosen intestinal cells. Supernatant was separated into a different 15ml falcon tube and this step was repeated three more times. A total of 4 supernatants were collected, namely epithelial fractions. 10µl of each fraction was used to evaluate their cellular content at the microscope. The last fractions are higher enriched in crypt/stem cells compared to the first collected fractions, and these were the ones usually used to plate. Chosen fractions were combined in a 50ml falcon tube and filled up with supplemented Advanced DMEM/F-12 media (ADF, Gibco) (Table 2.7) to centrifuge for 3min at 300*g*. Supernatant was discarded and the pellet was passed through a 70µm cell strainer (Scientific laboratory supplies) with 10ml of media to obtain isolated crypts. Crypts were spun in a 15ml falcon tube at 700*g* for 10min. Pellet was carefully resuspended with Matrigel (Corning) and

a volume of 20µl was plated in a pre-warmed 24-well plate. Crypts were incubated at 37ºC for 5-10min to allow Matrigel solidification. 500µl of pre-warmed growth media was added per well. Details of media and growth factors are illustrated in Table 2.7.

2.5.1.1 Epithelial fractionations

To evaluate the expression of *Rac1b* across the crypt-villus axis, fractions obtained from intestine epithelial cells isolation were used. Wild type or RAC1b^{OE} mice acutely induced with tamoxifen for 5 days were used for the experiment. SI was harvested and epithelial fractions were obtained as detailed in the previous section. Collected fractionations were separately spun down at 300*g* for 10min. The pellet was washed once with cold PBS and cells were spun at 300*g* for 5min. RNA was extracted from epithelial cells as detailed in section 2.3.1.3.

2.5.2 Tumour adenoma isolation

Harvested adenoma from mice was cut in small pieces in a petri dish and transferred to a 15ml falcon tube with ADF+. After washing the tumour twice with ADF+, tumour was incubated with 2ml of digestion media (Table 2.7) for 60-90min at 37ºC. While incubated, tumour was pipetted vigorously every 15min to facilitate cell dissociation. When digestion was completed, tumour cells were passed through a 70µm cell strainer and 100µl of 1% BSA was added to stop digestion. Cells were spun down at 600*g* for 3min and the pellet was washed twice with ADF. Final pellet was resuspended in an appropriate volume of Matrigel (up to 500µl for a 24-well plate) and plated in a pre-warmed plate.

2.5.3 Routine organoid culture maintenance

Organoid media was renewed every 3 to 4 days. When organoids reached confluency, these were split, usually up to 7 days. Organoids were mechanically passaged by scraping the wells with a 5ml pipette to disintegrate the Matrigel drop. Media was recollected and spun at 300*g* for 3 min. Supernatant was discarded and pellet was washed with 5ml of ADF+, pipetting up and down vigorously several times. This step was repeated twice until Matrigel

was dissociated from the cellular pellet. To physically break the crypts, the pellet was resuspended using a P200 with 200µl of ADF+ and centrifuged at 600*g* for 3min. Supernatant was aspirated and the pellet resuspended with Matrigel in a pre-warmed 24-well plate. Growth media was added when Matrigel was solidified in the incubator.

To freeze an organoid cell line, cells were dissociated from Matrigel and incubated for 15min at 37ºC with the freezing media. After incubation, 100µl of DMSO were gently added and cells were transferred to a freezing vial. Isopropanol freezing step is required prior to N_2 storage.

To thaw an organoid cell line, frozen vial was incubated for 30 seconds at 37ºC in the water bath. Aliquot was mixed with 9ml ADF+ and two washes with ADF+ were performed. Cells were broken with a P200 and embedded in Matrigel as usual. Importantly, Y-27632 was added in the growth media (at a concentration of 10µM) and removed after 2 days.

Supplemented ADF (ADF+)	Growth media (50ml)
Advanced DMEM/F12 medium (Gibco)	50ml ADF+
5ml PEN/STREP (HGU)	500µl N2 (Gibco)
5ml L-glutamine (2mM, HGU)	1000µl B27 (Life technologies)
5ml HEPES (Invitrogen)	50 ng/mL EGF (Invitrogen)
	100 ng/mL Noggin (Peprotech)

Adenoma digestion media	Freezing media
1700µl ADF+	900µl growth media
200µl Collagenase II (1mg/ml) (Sigma)	10µl EGF
100µl Hyalurodinase (0.5mg/ml) (Sigma)	10µl Noggin
4µl Y-27632 (10µM)	2µl Y-27632 (10µM)

Table 2.7: Organoid media conditions and supplemented growth factors.

2.5.4 Clonogenicity assay

This assay measures the ability of a tumoral cell to initiate a clone. APC P53 and APC P53 RAC1b^{OE} tumour organoids were washed and dissociated from Matrigel as usual (section 2.5.3). Crypt cells were chemically digested into single cells using 500-700µl of StemPro Accutase solution (Life technologies) for 10min at 37ºC. Digestion was stopped with ADF+

and cells were passed through a 40µm cell strainer (Scientific laboratories supplies). Cells were centrifuged at 600*g* for 5min and resuspended with 500µl of PBS supplemented with 5% of FCS. For an accurate count of single cells, a total number of 30,000 living and single cells were sorted through the flow cytometer (performed by Elisabeth Freyer, MRC, University of Edinburgh). These were centrifuged at 600*g* for 5min and resuspended in 150µl of Matrigel. Drops of 5µl Matrigel containing 1,000 single cells were plated in a 24-well plate. Growth media was added as usual and formed clones were scored 4 days after.

Clonogenicity assay of APC and APC RAC1b^{KO} tumour organoids was performed by Dr Patrizia Cammareri. Single cells were scored using the Countess machine.

2.5.5 EGFR inhibition with PD153035 treatment

APC and APC RAC1b^{KO} organoids were split as fragments at a density of 100 fragments/µl of Matrigel, distributed in drops of 15µl each. Cell fragments were achieved by gently breaking crypt cells with the P1000 instead of the P200, avoiding dissociation as single cells. After 24 hours, media was replaced by treatment media. Two different experimental conditions existed: control (1% DMSO) or EGFR inhibitor group (1µM PD153035, Sigma, cat# SML0564). DMSO or inhibitor were diluted in normal organoid growth media.

After 20 hours of treatment, pictures were taken for organoid area measurement, assessed with ImageJ. Furthermore, cell viability assay was conducted using Resazurin. Resazurin (R&D systems) was added to the media at 10% and cells were incubated for 7 hours. Fluorescence was read at 560/590nm emission wavelength using Victor Multilabel Plate Reader (Perkin Elmer).

In parallel, assay measuring organoid cell growth with or without EGF supplemented in the media was conducted by Dr Kevin Myant. Organoid area was evaluated by ImageJ software.

2.5.6 Acute and time course induction with TGF-β treatment

APC$^{-/-}$ P53 and APC$^{-/-}$ P53 RAC1b^{OE} organoids were digested into fragments and plated at a confluency of 100 fragments/µl of Matrigel (as in section 2.5.5). After 3 days when organoids have reached confluency, media was replaced by growth media supplemented with 5ng/ml of TGF-β1 ligand (Peprotech, cat# 100-21-10) or nothing as control. Organoids were collected after 2 hours for analysis of early response TGF-β target genes and RNA or protein was extracted (see sections 2.3.1.3 or 2.3.5.3 for details). For time point analysis, organoids were collected after 2, 4, 6 and 24 hours of TGF-β treatment and protein was extracted (section 2.3.5.3).

2.5.7 TGF-β-induced cell death

Around 40 fragments per 4µl drop of Matrigel were plated from APC$^{-/-}$ P53 and APC$^{-/-}$ P53 RAC1b^{OE} organoids. The number of clones formed was scored after 48 hours and the growth media was replaced with TGF-β1-supplemented growth media. A titration of TGF-β1 doses was used to evaluate cell death: 0, 0.3ng/ml, 1ng/ml, 3ng/ml, 10ng/ml and 30ng/ml. Dead organoids were scored following 24 hours of treatment, using the work from Grabinger *et al.* as a reference for the morphology of dead organoids[439]. Percentage of cell death was calculated individually per well compared to the initial number of organoids.

Furthermore, cell death was measured using the CellEvent™ Caspase-3/7 Green Detection Reagent (Invitrogen). After 24 hours that organoids were split as fragments, 1000X of reagent was added to the media, which would release fluorescent light upon reagent substrate cleavage by caspase-3 and -7 activation. Both types of images, brightfield and fluorescent, were taken following 24 hours of treatment. Quantification was carried out on ImageJ, whereby brightfield pictures were used as a reference for organoid ROIs (region of interest) to overlay with fluorescent images and to calculate the percentage of green pixels per organoid. Validation of this assay was conducted by simultaneous scoring of cell death according to organoid morphology.

2.6 TCGA dataset analysis

Synapse is a database system to allow public access to the data generated by the TCGA program, as well as by the Colorectal Cancer Subtyping Consortium (CRCSC).

Isoform data from TCGA COAD tumours were downloaded from the TCGA Synapse group (synapse accession: doi:10.7303/syn300013)[440] and "IlluminaHiSeq_RNASeqV2" study. Expression of *RAC1* (uc003spx.3) and *RAC1B* (uc003spw.3) was compared in tumours and in matched normal tissue.

PSI data was downloaded from the TCGASpliceSeq web-based platform[441]. Data for *RAC1* PSI was downloaded per each tumour type. Types which had normal tissue expression were compared with tumour expression. Moreover, COAD tumours were subdivided according to their *RAC1* PSI:

- $RAC1B^{high}$: if PSI value > 0.2 above the average PSI of normal tissue
- $RAC1B^{low}$: if PSI value < average PSI of normal tissue

Clinical data from TCGA COAD tumours was extracted from the CBioPortal website, which included survival, lymphovascular invasion and metastasis data. TCGA tumour codes were overlapped with those from the TCGASpliceSeq and survival and invasion data were classified according to $RAC1B^{high}$ or $RAC1B^{low}$ groups.

Pathway analysis was also facilitated by the CBioPortal website, whereby TCGA tumours can be analysed according to the presence or absence of mutations in a particular gene.

RSEM-calculated (RNA-Seq by Expectation Maximization) reads of COAD tumours were downloaded from the analysis performed by Wang *et al*[442]. Based on the CMS gene signatures (provided to the laboratory), RSEM of TCGA tumours were distributed according to the CMS gene profile. PSI values were associated to their corresponding TCGA code and both, PSI and RSEM values were correlated by a Pearson's test. Moreover, CMS groups were compared each other according to *RAC1* PSI.

Expression of particular genes (such as *LRG5* or *AXIN2*) was calculated from the RSEM data set overlapped with *RAC1* PSI values.

For CMS survival analyses, TCGA codes with CMS and *RAC1* PSI information were compared to the survival information downloaded from the CBioPortal clinical data.

Tumour stage data was acquired from the clinical data published by the CRCSC synapse group (accession: doi: 10.7303/syn2623706), whose TCGA codes were overlapped with the corresponding *RAC1* PSI values. Moreover, MSI status and *BRAF* mutations were also facilitated by this dataset.

Finally, RPPA data was downloaded from The Cancer Proteome Atlas (https://tcpaportal.org/tcpa/download.html). Protein expression values were correlated with *RAC1* PSI by Pearson's correlation.

2.7 Material list from IGMM technical services

	Solution	Ingredients	Amount/litre	Volume
Lab solutions	50X TAE	Tris Base	242g	500ml
		Glacial Acetic acid	57.1ml	
	0.5M EDTA	EDTA	186.1g	500ml
		NaOH to pH	~20.0g	
Bug media	L-Broth	Tryptone	10.0g	400ml
		Yeast extract	5.0g	
		Sodium Chloride	10.0g	
		Glucose	1.0g	
Tissue culture	FCS			20ml
	Dulbeccos	Unipath (Oxoid) Cat Nos. BR149		500ml
	L-Glutamine	L-Glutamine	30.0g	20ml
	Penicillin/	Penicillin	7.0g ($10x10^8$u)	20ml
	Streptomycin	Streptomycin	13.0g	

Table 2.8: List of solutions and ingredients used in this thesis produced by the IGMM technical services.

2.8 Statistical analysis

GraphPad Prism version v7.0 software was used for all statistical analyses. Data was analysed as follows:

- Comparison between two groups was assessed by Student's T-test, 2-tailed, parametric and unpaired. If groups did not follow normality, a non-parametric Mann-Whitney test was conducted instead.
- Percentage of survival was analysed by a Log-rank (Mantel-Cox) test.
- Correlation between two data sets was evaluated by 2-tailed Pearson correlation, with a 95% of confidence interval.
- Association between two categories was assessed by two-sided Fisher's test

P values < 0.05 were considered statistically significant. Levels of significance are labelled as follows: * = $p < 0.05$; ** = $p < 0.001$; *** = $p < 0.001$; **** $p \leq 0.0001$. Not significant results are labelled as n.s. ($p > 0.05$).

Chapter 3: Characterization of Rac1b expression

3.1 Localization of Rac1b in the mouse intestine

Little is known about where Rac1b is localized within a tissue. To date, most investigations of Rac1b expression have been made in 2D cell lines, so its subcellular localization is better defined than its tissue distribution. Given that it is a small GTPase protein, and its best characterised for its GTP bound state, it is expected to be membrane bound as Rac1 is in its active version[443]. Early studies investigating Rac1b function showed that it is predominantly localised at the membrane, where is functionally active[375]. For instance, Lee et al., demonstrated that Rac1b-dependent ROS production only occurs when Rac1b is retained in the membrane with fibronectin[393,394]. IHC staining for Rac1b in breast cancer biopsies validated a membrane localisation of Rac1b in the most malignant tumours, which potentially would also present ROS-driven EMT activation. Other investigators have also shown immunohistochemical analysis from human cancer tissue[395,400] and all of them have used the unique commercial Rac1b antibody available except Li *et al*., who generated their own[380]. Indeed, they were pioneers in publishing an *in vivo* investigation deleting *Rac1b*. Initially, they studied Rac1b function *in vitro* using transduced colonic cells with lentivirus overexpressing either *Rac1* or *Rac1b* and saw that, while Rac1 was mainly localized in the nucleus or the cytoplasm, Rac1b had a membrane and cytoplasmatic localization[380]. Then, they knocked out *Rac1b* using CRIPSR, only affecting the exon 3b and without altering *Rac1* expression. They showed by immunohistochemistry that, in normal conditions, Rac1b was expressed in the colonic epithelium. This is validated by their knock-out rats, which are completely negative for Rac1b staining, although the low-resolution images used in the paper make a definite conclusion in tissue location difficult. Regarding murine models, to date there is no *in vivo* data describing Rac1b localization in either normal tissue or tumours. Recently, Chastre *et al.* presented their *in vivo* work using a Villin promoter and the Cre recombinase system to overexpress *Rac1b* in an $Apc^{Min/+}$ mice model[409], but they only characterized Rac1b overexpression by Western blot, leaving questions regarding Rac1b localization in the murine intestine.

3.1.1 Unespecific signal with Rac1b antibody

The aim of this project is to investigate the function of Rac1b in both normal and tumour intestine tissue by modulating its expression in mouse models. Therefore, it is important to characterise its tissue location for a better functional description. To this end, I harvested intestinal tissue from wild type (WT) mice and fixed in PFA for paraffin wax embedding and microtome sectioning. Since I also have available mice that either overexpress a copy of *Rac1b* in the Rosa26 locus ($RAC1B^{OE}$) or that have deleted the exon 3b ($RAC1B^{KO}$) specifically in the intestine (more detail of these models explained further in this chapter), intestines from these mice were also harvested as positive and negative controls, respectively. As mentioned above, there is only a commercial Rac1b antibody available from Millipore which I used for immunohistochemistry (IHC). After several IHC protocol optimizations, I observed an interesting pattern of staining in the WT mice, with a punctuated cytoplasmatic staining in the crypts, which looked very alike the normal breast biopsies from Lee *et al.* paper[393], and a membrane staining as it goes up the villus axis (Figure 1A). However, performing the same IHC protocol in the control sections showed an identical pattern and intensity of staining (Figure 1B and C), indicating that the antibody is not binding specifically to Rac1b antigen. Of note, $RAC1B^{OE}$ mice overexpress a human *Rac1b* cDNA, implying that the antibody recognises neither mouse nor human *Rac1b*. Thence, I was unable to validate this antibody for IHC and, from this perspective, it potentially calls into question the data presented in the literature using it, especially the studies which have used this antibody to stain cancer tissue microarray (TMA) sections.

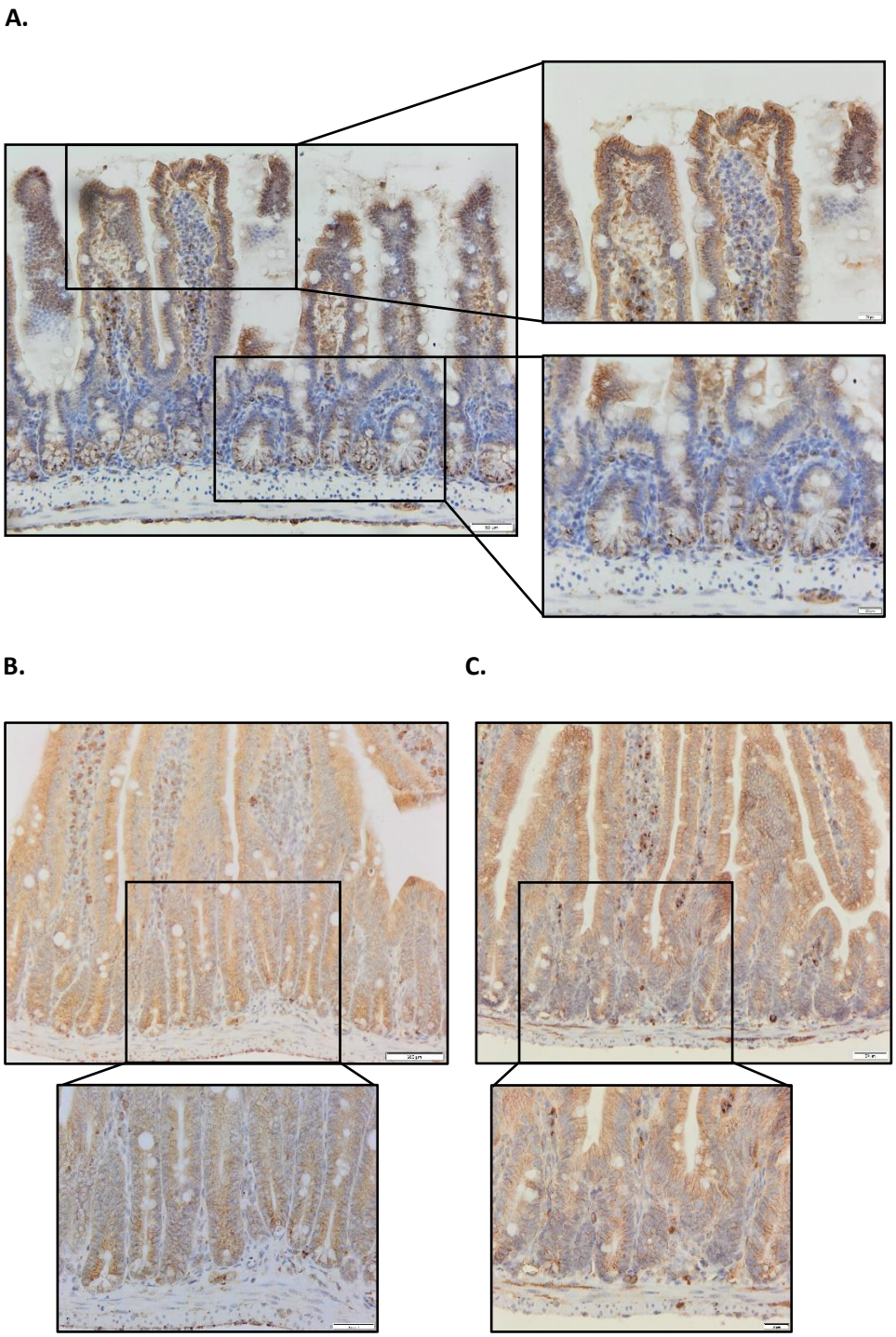


Figure 3.1: Rac1b immunohistochemistry of the murine intestine. Images for tissue staining with Rac1b antibody in WT (A), $RAC1B^{OE}$ (B) and $RAC1B^{KO}$ (C) mice. Magnification at 20x.

3.1.2 Intestinal epithelial fractionation for *Rac1b* characterization

Hence, I sought an alternative method to study Rac1b localization in the small intestine *in vivo* but at the RNA level. To do so, I made use of an intestinal epithelial extraction protocol to fractionate the epithelium into different grades of crypt enrichment, subsequent extraction of RNA from each fraction and then, cDNA synthesis for qRT-PCR analysis. In brief, extraction of intestinal epithelial cells consists on the scraping of the inner part of the intestine to get rid of the villus and thorough sequential washes of PBS:EDTA to collect released intestinal epithelial cells, with each wash corresponding to a fraction more enriched for cells closer to the crypt base. An outline of each fraction is schematised in figure 2A: the last fraction extracted (3) represents the most crypt-enriched part of the intestine, presumably containing the stem cells, while the villus would be highly enriched with intestinal differentiation markers. As a characterization of the fractions, I investigated the expression of *EphB2* and *Krt20*, a crypt and a differentiation marker respectively[444]. As can be seen in figure 2C, their expression changes reciprocally as fractions become more crypt-enriched, with *EphB2* being more highly expressed in fraction 3 and *Krt20* significantly decreased at the last fraction. For a further validation, I also looked at the expression of stem cell markers such as *Lgr5*[131] and *Scl14a1*[445], which are almost uniquely expressed in fraction 2 and 3, and at the expression of proliferation markers like *c-Myc*[446] and *Tiam1*[345], which have an increased gradient of expression towards the last fractions but not exclusively (Figure 2B and D). Once fractions were validated to be representative of a crypt-villus axis, I looked at the expression of *Rac1* and *Rac1b*. *Rac1* presents a fairly flat pattern of expression across the different fractions, only being slightly increased in fraction 3 (Figure 2E). This is consistent with previous studies demonstrating control of Rac1 activity via the expression of regulatory genes[447]. Indeed, as seen in the previous figure, *Tiam1*, a Rac1-GEF, has an increased gradation of expression towards the most crypt-enriched fractions, presumably maintaining a high crypt Rac1 activity as has been previously demonstrated by Myant and collegues[292]. In contrast, *Rac1b* expression significantly increases in each fraction, reaching its highest expression at fraction 3 (Figure 2F). However, there is a considerable amount of *Rac1b* transcript in the other fractions, so *Rac1b* cannot be considered as a stem cell marker. Instead, it behaves more like *c-Myc* or *Tiam1*, increasing gradually from the villus down to the crypts. Therefore, these data suggest that *Rac1b* transcript is expressed with elevated expression in the proliferative crypts

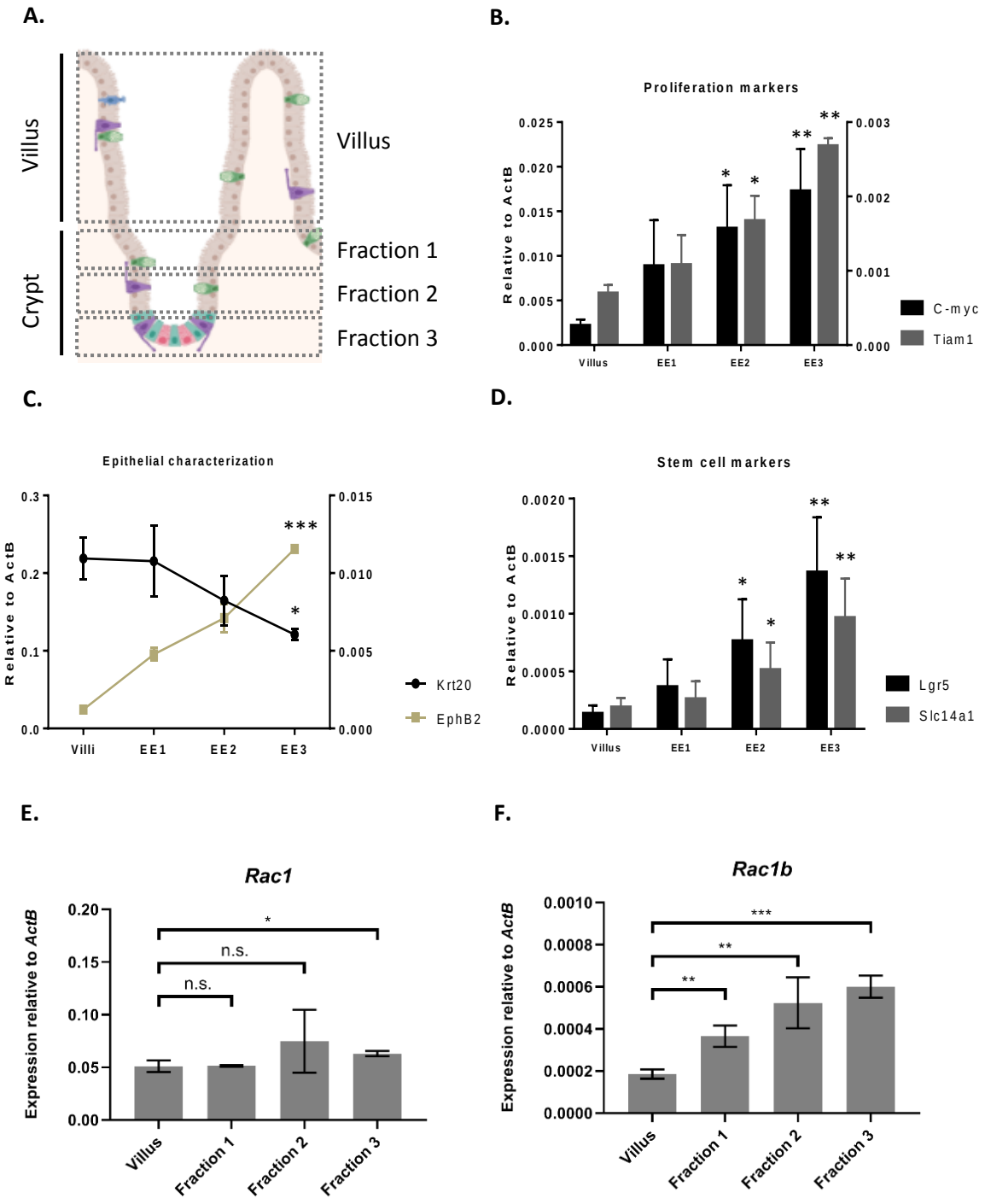


Figure 3.2: Intestinal fractionation experiment characterises a crypt-base location of Rac1b. A: Schematic representation of each epithelial extraction (EE) fraction. B, C and D: Validation of the epithelial fractions by the expression of differentiated (*Krt20*), crypt-base (*EphB2*), proliferative (*c-myc* and *Tiam1*) and stem cell markers (*Lgr5* and *Slc14a1*). E: Expression of *Rac1* is consistent across fractions. F: *Rac1b* is higher expressed at the intestinal proliferative zone. T-test: *= p≤0.05, **= p≤0.01, ***= p≤0.001, n.s. = not significant, p>0.05. N = 3 vs 3.

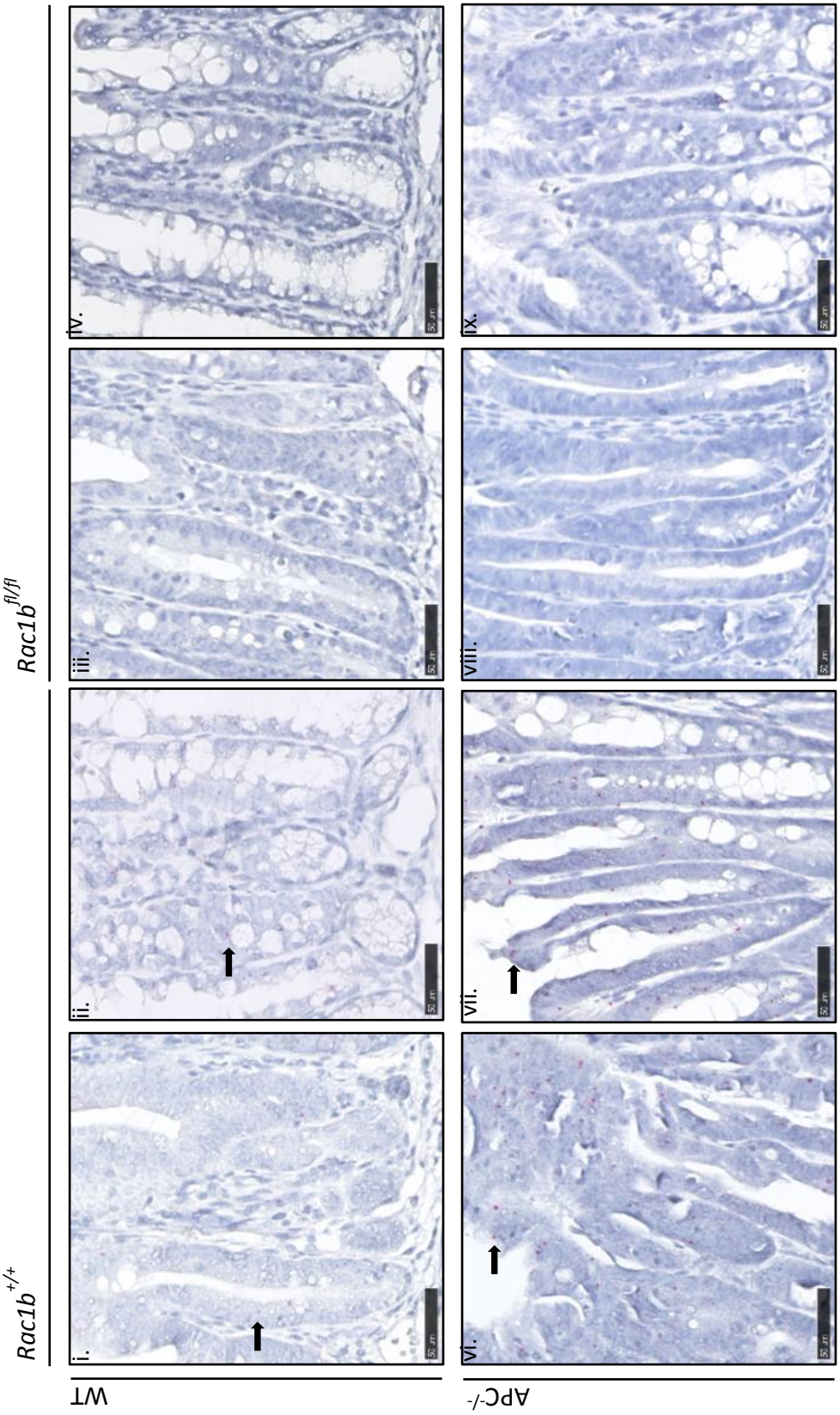
3.1.3 Designing a BaseScope probe for exon 3b

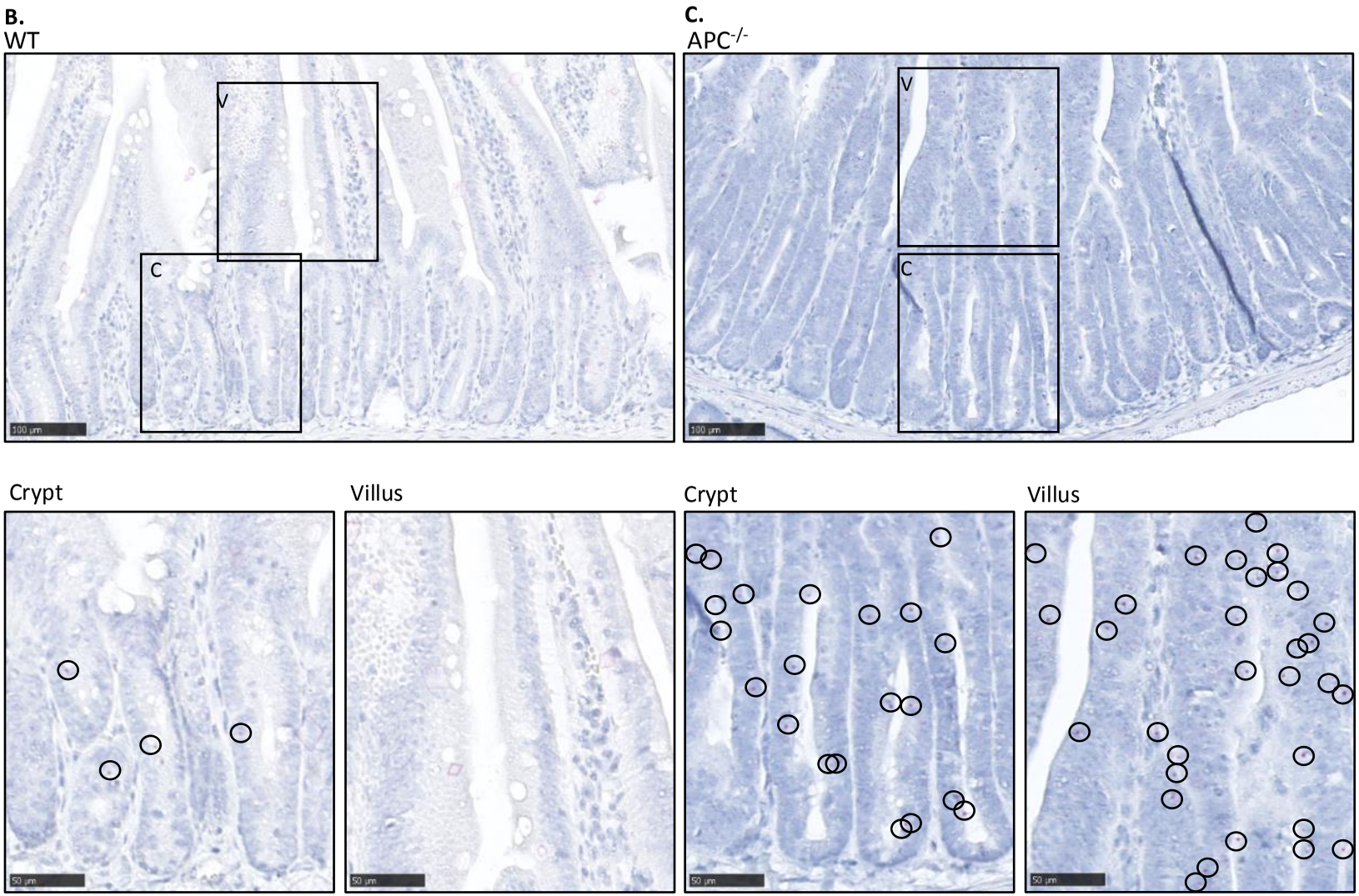
Even though this fractionation experiment provided important information about Rac1b location and potential role, I wanted to visualise Rac1b expression in the context of whole tissue. To this end, I used an RNA *in situ* hybridization method named BaseScope, specifically designed to discern between closely related isoforms. This technique, as RNAscope, consists of a probe that specifically hybridises to a targeted RNA and provides a signal that will be amplified by a series of reagents. The advantage BaseScope offers among RNAscope is that it can detect short RNA sequences, difficult RNA targets or, as in the current interest, detection of exon junctions, being ideal to differentiate between splice isoforms. We designed a probe that specifically bound the exon 3-3b junction, which corresponds to the end of the exon 3 and the beginning of the following exon (Figure 2.2). Therefore, this probe would only release signal when it hybridises to *Rac1b*-exon3b containing isoform and not *Rac1*. We stained wild type mice small intestine (SI) and large intestine (LI) sections and $RAC1B^{KO}$ as a negative control with the designed probe. Signal is observed as small pink dots every time the probe detects a *Rac1b* transcript molecule. As shown in figure 3.3A, an average of 2-3 dots per crypt are detected in the SI and it is significantly more abundant in the LI than in the SI (Figure 3.3D). Sections were also stained with the PPIB (positive) and DAPB (negative) probes for RNA quality control of the tissue (Appendix 1). Importantly, BaseScope staining also reveals a crypt preferential Rac1b localization, with significantly more transcript-dots in the crypts compared to the villus, reaffirming the crypt position of Rac1b seen with the fractionation experiment. Moreover, the $RAC1B^{KO}$ negative control presents virtually no staining at all, validating both the technique and my KO model.

Mutations in the tumour suppressor gene *Adenomatous polyposis coli* (*Apc*) is a key event in the vast majority of sporadic human CRC and the main characteristic for the low frequency hereditary Familial Adenomatous Polyposis (FAP) syndrome[50]. Mouse models for *in vivo* ageing modelling of CRC usually harbour heterozygous mutations in the *Apc* gene, which do not present a phenotype until the remaining copy of *Apc* is lost. On the contrary, mice with homozygous deletion of *Apc* cannot be used for long-term experiments, since the complete loss of *Apc* causes a rapid nuclear translocation of β-catenin, activation of the WNT signalling pathway, misslocalization of Paneth and stem cells and a consequent intestinal hyperproliferation after 5 days of tamoxifen induction[414]. However, this phenotype corresponds to early CRC lesions and it has become a suitable model to study acute

activations of the WNT pathway along with other mutations in a cell population relatively homogenous without compensatory mechanisms. Therefore, I sought to determine whether *Rac1b* transcript localization would be altered immediately after *Apc* loss. By using a tamoxifen inducible Cre recombinase that targets specifically intestinal and colonic cells by a Villin promoter (*VilCreER*^{T2}), the *Apc* gene was targeted for deletion. Mice with homozygous *Apc* deletion (*VilCre*^{ERT2}*Apc*$^{fl/fl}$*Rac1b*$^{+/+}$, from now on APC$^{-/-}$) and their Rac1b controls (*VilCre*^{ERT2}*Apc*$^{fl/fl}$*Rac1b*$^{fl/fl}$, hereafter named APC$^{-/-}$ RAC1B^{KO}) were induced for 5 days and tissue was fixed and sectioned for BaseScope staining. Intriguingly, Rac1b^{+ve} signal dramatically increased compared to wild type mice (Figure 3.3F). In the SI, it still maintained its preferential crypt position, but it was not as delimited as before due to abnormal intestinal proliferation and crypts hyperplasia (Figure 3.3E). This result is confirmed by qRT-PCR, where *Rac1b* expression is significantly increased in *Apc* deficient tissue compared to WT (Figure 3.3G). In conclusion, BaseScope staining demonstrates *Rac1b* to be mainly localised in the crypts, that it is more abundant in the LI compared to the SI and that its expression increases after the loss of both copies of the *Apc* gene.

A.




B.
WT
V
C
100 µm
C.
APC-/-
V
C
100 µm
Crypt
Villus
Crypt
Villus
50 µm
50 µm
50 µm
50 µm

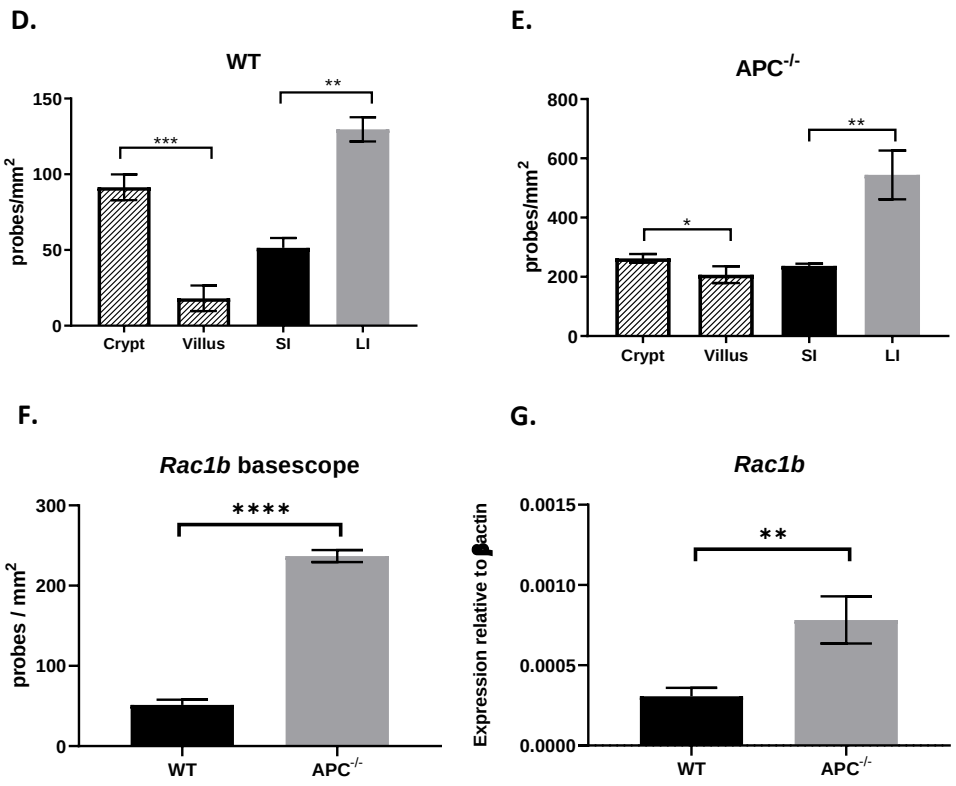


Figure 3.3: Visualisation of *Rac1b* transcript in the murine intestine by BaseScope. A: an average of 2 to 3 dots are detected in the SI (i) and LI (ii) of WT mice, while its expression significantly raises upon *Apc* deletion (vi and vii). Probe specificity is validated by staining of *Rac1b*-deficient mice (iii and iv for SI and LI of WT mice respectively; viii and ix for SI and LI of $APC^{-/-}$ mice respectively). Arrows mark an example of a dot in each positive image. B and C show an overview of the SI in WT and $APC^{-/-}$ mice, respectively. Amplification of crypt and villus regions are illustrated below; circles highlight positive signal (pink dots).
Probe scores in WT (D) and $APC^{-/-}$ (E) tissue, comparing crypt vs villus and SI vs LI. F: total number of probes scores in WT and $APC^{-/-}$ tissue. G: qRT-PCR for *Rac1b* in WT and $APC^{-/-}$, validating the increased expression in *Apc*-deficient tissue observed by BaseScope.
T-test: *= p≤0.05, **= p≤0.01, ***= p≤0.001, ****= p≤0.0001. N = 3 vs 3.
BaseScope staining conducted in collaboration with Colin Nixon.

3.1.4 Summary

Colorectal cancer is one of the diseases Rac1b has been most investigated in, together with breast and lung cancer. However, most of the studies focused their effort on its functional characterization, leaving a gap in other aspects of its characterization such as cellular and tissue localization. Besides, since it has been mainly studied in 2D cell lines, there is a lack of *in vivo* characterization and validation of the current knowledge. In this section I aimed to depict Rac1b localization in mice intestine using wild type and genetic modified mouse models with either *Rac1b* overexpression or deletion. I initially tested a commercial Rac1b antibody in the murine intestine of WT mice. After several IHC optimization and using as reference the positive ($RAC1b^{OE}$) and negative ($RAC1b^{KO}$) mice controls, I was unable to validate its specificity so I sought an alternative methodology to study its intestinal expression. Using a fractionation experiment and BaseScope staining, we found that *Rac1b* is mainly expressed in the crypt zone. It presents a low expression in normal conditions, which is more abundant in the LI than in the SI, and expression significantly increases after *Apc* deletion, reaffirming its potential tumour-related role.

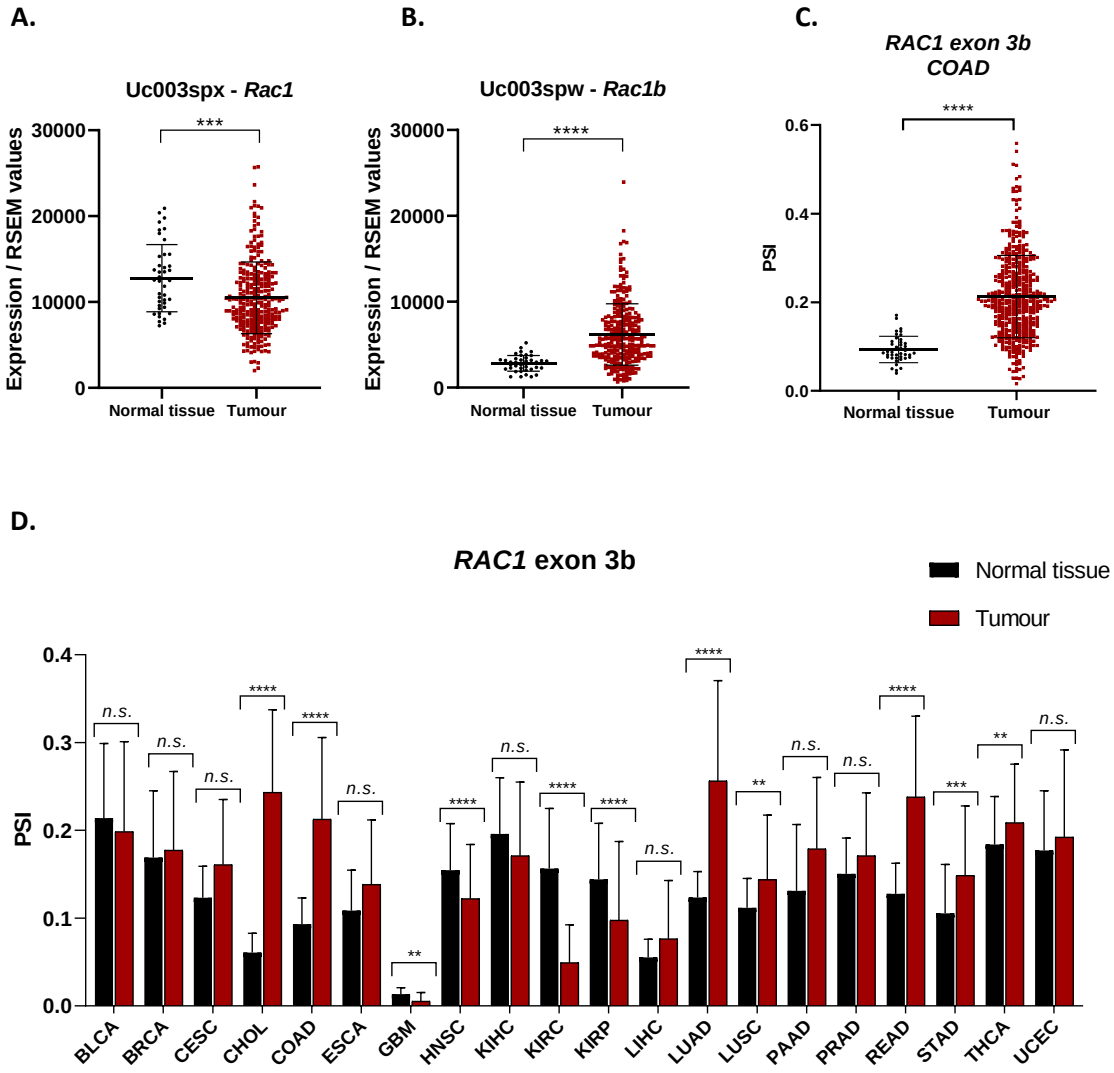
3.2 *RAC1B* expression in human CRC

3.2.1 *RAC1B* as a tumour-related isoform

Rac1b was first discovered in tumoral 2D cells[372,448] and since then it had been attributed a tumour promoter role. However, further investigations are needed to confer a more direct human relevance in cancer. In the last couple of decades, advances in genome sequencing techniques and bioinformatic analysis have made it possible to explore at an *in-silico* scale diseases such as cancer, and facilitate the discovery of both tumour suppressor and oncogenes by differentially expressed genes (DEGs) between tumoral and normal tissue. One of the key tools for cancer researchers from the bioinformatic point of view was the realisation of The Cancer Genome Atlas (TCGA). Data generated from the TCGA is stored in Sage Bionetworks' Synapse database and is available for public download. For my project, I made use of this source to extract isoform expression data for both *Rac1* and *Rac1b* in colorectal cancer (COAD). By downloading the data from Synapse and splitting normal and tumour samples I saw that the estimated counts for *Uc003spx* (*Rac1* isoform) were significantly lower in tumours compared to normal tissue (Figure 3.4A, p=0.0005). On the contrary, *Uc003spw* (*Rac1b* isoform) is strongly significantly elevated in tumours (Figure 3.4B, p<0.0001). Given that Rac1 signalling is required and activated during CRC tumorigenesis[292] but its expression diminishes in tumours (Figure 3.4A), it highly implies alternative methods for pathway activation, such as the observed increase in the *Rac1b* isoform.

Aside from the raw data downloaded from Synapse, other groups have made efforts to create comprehensive and user friendly web platforms to help with the general analysis of a specific target gene or cancer type, such as cBioPortal[449,450] or Xena[451]. Given that *Rac1b* is an alternative splice isoform, other web bioinformatic platforms able to discern between similar isoforms are needed. Ryan *et al.* created a platform specifically designed for the study of alternative spliced genes in cancer using the TCGA database, named TCGA SpliceSeq[441]. They previously designed SpliceSeq, a Java application that allows RNAseq transcript expression visualization through splice graphs and further functional prediction by protein sequence translation[452]. The TCGA isoform expression data was run through SpliceSeq to obtain a complete and comparative alternative splice analysis of specific genes across different tumour types which is stored in a public web-based interactive database, the TCGA SpliceSeq. A gene of interest or the top genes with most significant splicing events can be

selected for one or more tumour types and among the different results of the search, appears a splice graph and the Uniprot sequence, a short description, the type/s of splicing event/s and a Percent Spliced In (PSI) graph comparing tumour vs normal tissue. The PSI measures the number of transcript reads that contain an exon and is a value used for quantifying splicing events[361]. Indeed, in the paper presenting the TCGA SpliceSeq, *Rac1* was used as a queried gene example to explore its splicing events and showed that the isoform with exon 4 inclusion (*Rac1b*) has higher PSI in tumours compared to normal tissue. Prompted by this result, I used the TCGA SpliceSeq website to download the PSIs values of *Rac1* for each tumour type that contained matching normal tissue data and conducted statistical analysis. As shown in figure 3.4D, cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD) and thyroid carcinoma (THCA) presented significantly increased levels of *Rac1* PSI in tumours compared to normal tissue. On the contrary, glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC) and kidney renal papillary cell carcinoma (KIRP) showed the opposite trend, with *Rac1b* underrepresented in tumours according to its PSI. For a more general concern, it would be interesting to study similarities/shared properties between these two groups of tumour types (increased vs decreased *Rac1* PSI). Several studies have been made using TCGA data to cluster different cancers according to its mutations, as the one published by Kandoth *et al.* in 2013. In their work, albeit they only analysed 12 tumour types, it can be seen that COAD/READ, LUAD and LUSC have the highest percentage of significantly mutated genes in WNT/β-catenin, MAPK and TGF-β signalling compared to GBM, HNSC and KIRC, which have the lowest[83]. Even though this could lead to an intriguing literature and TCGA bioinformatic research, it was not the aim of the current project and I kept focus on the main finding from the TCGA SpliceSeq analysis: COAD was one of the cancer types with the most significant increase in *RAC1B* splicing (Figure 3.4C, $p \leq 0.0001$).



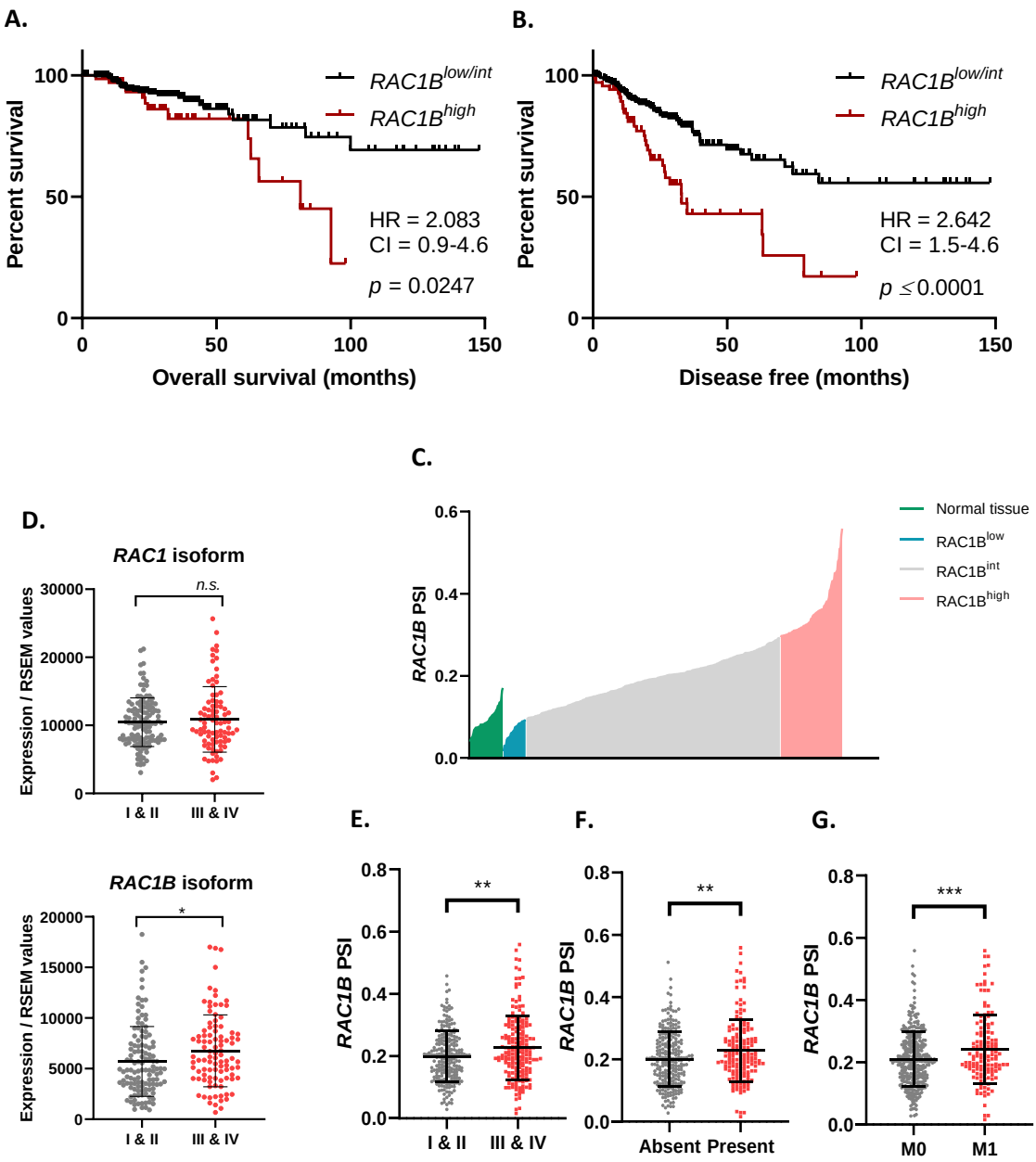
<u>Figure 3.4</u>: Expression of *RAC1B* increases in human cancers. A and B: RSEM counts for *Uc003spc/Rac1* (A) and *Uc003spw/Rac1b* (B) isoforms comparing colorectal tumours and matching normal tissue. C: PSI values for Rac1 exon inclusion (*RAC1B*) in COAD TCGA dataset for normal tissue and tumour. D: PSI values for Rac1b across 20 different tumour types comparing normal tissue and tumour.

Mann-Whitney test: *= p≤0.05, **= p≤0.01, ***= p≤0.001, ****= p≤0.0001, n.s. = not significant, p>0.05.

PSI: Percent Spliced In, BLCA: bladder urothelial carcinoma, BRCA: breast invasive carcinoma, CESC: cervical squamous cell carcinoma and endocervical adenocarcinoma, CHOL: cholangiocarcinoma, COAD: colon adenocarcinoma, ESCA: oesophageal carcinoma, GBM: glioblastoma multiforme, HNSC: head and neck squamous cell carcinoma, KIHC: kidney chromophobe, KIRC: kidney renal clear cell carcinoma, KIRP: kidney renal papillary cell carcinoma, LIHC: liver hepatocellular carcinoma, LUAD: lung adenocarcinoma, LUSC: lung squamous cell carcinoma, PAAD: pancreatic adenocarcinoma, PRAD: prostate adenocarcinoma, READ: rectum adenocarcinoma, STAD: stomach adenocarcinoma, THCA: thyroid carcinoma, UCEC: uterine corpus endometrial carcinoma.

Individual *RAC1B* PSI values with TCGA tumour identifiers were downloaded and overlapped with TCGA clinical data, such as patient survival or tumour stage. First, I classified tumours as high or low for *Rac1* PSI using thresholds taken from Kahles *et al.*[371]. Using a more stringent threshold a Rac1b high group ($RAC1B^{high}$) was designated if the PSI value was higher than 0.2 delta PSI of normal tissue. Contrariwise a Rac1b low group ($RAC1B^{low}$) was designated when the PSI value was lower than the average of normal tissue PSI, leaving a group of Rac1b intermediate expression group ($RAC1B^{int}$) (Figure 3.5C). Overall survival (time between disease diagnosis and patient alive status) and the disease-free survival (time between treatment and signs of disease) was analysed. I compared both between $RAC1B^{high}$ and $RAC1B^{int+low}$ (all tumours that are not classified as high) groups by overlapping TCGA codes. Interestingly, patients with $RAC1B^{high}$ tumours have a significantly poorer overall survival and disease-free survival compared to patients with $RAC1B^{int+low}$ tumours (Figure 3.5A and B; p=0.0247 and p≤0.0001 respectively). Therefore, high *RAC1B* levels in CRC correlate with worse survival prognosis.

Seeing that Rac1b might confer an advantage to tumours compared to non-tumoral tissue, I wondered whether *RAC1B* expression correlates with advanced tumour stages. Clinical data downloaded from CbioPortal using COAD TCGA study data set provided information about tumour stage and its invasion into the lymphovascular system. Interestingly, tumours with the highest PSI value for *RAC1B* were enriched for invasive stages (III and IV) compared to non-invasive (p=0.0029, figure 3.5E). The same analysis was carried out using the isoform data from figure 3.4A and B and overlapping with CbioPortal TCGA codes, showing that only *RAC1B* isoform but not *RAC1* correlated with stages III and IV (p=0.0283, Figure 3.5E). Consequently, tumours with high *RAC1B* were more likely to present with invasion into the lymphovascular system (p=0.004, Figure 3.5F) and metastasis (p=0.0008, Figure 3.5G). Altogether, these analyses indicate that Rac1b is preferentially expressed in tumours compared to normal tissue, and correlates with malignant behaviour, driving invasion into blood vessels and the lymphatic system and progression to metastatic disease, traits of advanced CRC stages.



<u>Figure 3.5</u>: Expression of *RAC1B* in CRC correlates with poor prognosis and tumour invasion. A and B: Kaplan-Meier plots for overall survival (A) and disease-free survival (B) for COAD patients from TCGA dataset (Log-rank test). C: Distribution of COAD TCGA normal tissue and tumours based on *RAC1B* PSI values. D: RSEM counts for *RAC1* and *RAC1B* isoforms comparing expression in early (I&II) or late (III & IV) stages. E, F and G: Evaluation of tumour stages (E), presence or absence of lymphovascular invasion (F) and metastatic disease (G) according to *RAC1B* PSI values.
M0: no distant metastasis, M1: metastasis. HR: Hazard Ratio, CI: Confidence Interval.
Mann-Whitney test: *= p≤0.05, **= p≤0.01, ***= p≤0.001, ****= p≤0.0001, n.s. = not significant, p>0.05.

3.2.2 CMS tumour stratification and correlations based on *RAC1B* expression

The TCGA data has facilitated the molecular stratification of the majority of tumour types analysed[197,198]. Accurate tumour classifications allow more precise information for clinical decision making. Compared to previous CRC classification merely based on their MSS status (either MSI or MSS), recent joint efforts from the CRC subtyping Consortium group (CRCSC) identified four RNA-based groups dubbed the CMSs subtypes[185]. Briefly, the CMS include the immune-related CMS1, the canonical WNT/β-catenin-activated CMS2, the metabolic rearranged CMS3 and the mesenchymal/TGF-β enriched CMS4. Since they were published and accepted as a "gold standard" classification method in CRC, assigning a CMS label to a certain tumour becomes of an indubitable need, especially when it comes to treatment and survival likelihood prediction for the patient. After revealing the worse prognosis for patients with high *RAC1B* levels (Figure 3.5), I sought to determine whether there was a correlation between these tumours and any of the CMS subtypes. To do so, I downloaded the RSEM expected counts data from the TCGA RNAseq analysis within the COAD data set[442] and extracted expression data of the genes that classify the different CMS subtypes and correlated their expression to *RAC1B* PSI values (list of CMS genes provided from a colleague to the lab). As illustrated in Figure 3.6A, few genes that classify CMS3 and CMS4 correlated with *RAC1B* PSI. However, CMS1 and CMS2 showed interesting correlations with *RAC1B* expression. In the case of CMS1, the majority of genes that classify this subtype negatively correlated with *RAC1B* expression (66% of the signature genes significantly negative correlated (blue zone)). In contrast, almost 53% of the genes that classify CMS2 had a positive correlation with *RAC1B* (red zone). Given that the CRCSC uploaded in Synapse the clinical data from the TCGA data set they used, and each tumour is labelled with the CMS type that was assigned, I could also demonstrate these correlations by overlapping these TCGA codes with the TCGA *RAC1B* PSI codes and plot *RAC1B* PSI values according to each subtype, demonstrating the increased *RAC1B* expression in CMS2 tumours and reduced expression in CMS1 tumours (Figure 3.6B). Moreover, and as mentioned previously, there are certain characteristics that define each CMSs type, like the TGF-β signature in the CMS4 or the WNT pathway activation in the CMS2. In order to fully support the observed correlations, I sought whether these characteristics were also in line with Rac1-exon 3b insertion. Increased expression of the targeted genes of a signalling pathway is generally accepted as a read out for pathway activation. *LGR5*, *AXIN2*, *ASCL2* and *TCF7* are canonical target genes of the WNT/β-Catenin signalling pathway and their RSEM values are all highly

significantly increased in the RAC1B^{high} compared to the RAC1B^{low}, validating the positive correlation with the CMS2 in tumours with higher *Rac1b* expression (Figure 3.7A). On the contrary, CMS1 is defined for its high percentage of MSI tumours and *BRAF* mutation. The same analysis revealed reduced expressing of *RAC1B* in tumours positive for MSI and/or *B-RAF* mutation, as hypothesised by the negative correlation with CMS1 (Figure 3.7B). This data was also validated using the TSVdb software, another TCGA web-based platform specialised in alternative spliced isoform analysis[453] (Figure 3.7C).

Another useful parameter one could extract from the clinical data provided by the CRCSC was the disease-free survival per each tumour labelled with a CMS type. As an additional validation of the CMSs prognosis presented in the paper, survival data for each subtype is plotted in figure 3.8A. CMS4 has the worst relapse-free survival among the four subtypes and CMS2 one of the best prognosis, along with CMS3. Interestingly, I saw in the previous analysis that high levels of *RAC1B* predispose patients to a worse overall survival and disease-free survival (Figure 3.5). This is in apparent contradiction to my finding that *RAC1B* levels in tumours correlate with the relatively good prognosis CMS2 subtype (Figure 3.8B). To further investigate this, I analysed disease-free survival in patients with CMS2 tumours according to *RAC1B* expression (keeping the same RAC1B^{high}/RAC1B$^{low+int}$ grouping methodology as before). Interestingly, patients from the CMS2 group with high *RAC1B* expression had very poor prognosis, with a similar survival outcome as patients with CMS4' tumours. Therefore, this apparent contradiction is resolved if tumours within the CMS2 subtype are secondarily categorised according to *RAC1B* levels, demonstrating that patients with the worst survival were potentially driven by high *RAC1B* expression. Altogether, these results would suggest that *RAC1B* expression is higher in the WNT activated CMS2 subtype and characterizing *RAC1B* levels within a tumour might help to define survival prognosis. Besides, defining its strong correlation with CMS2 and antagonism with CMS1 may also shed light on its function in cancer.

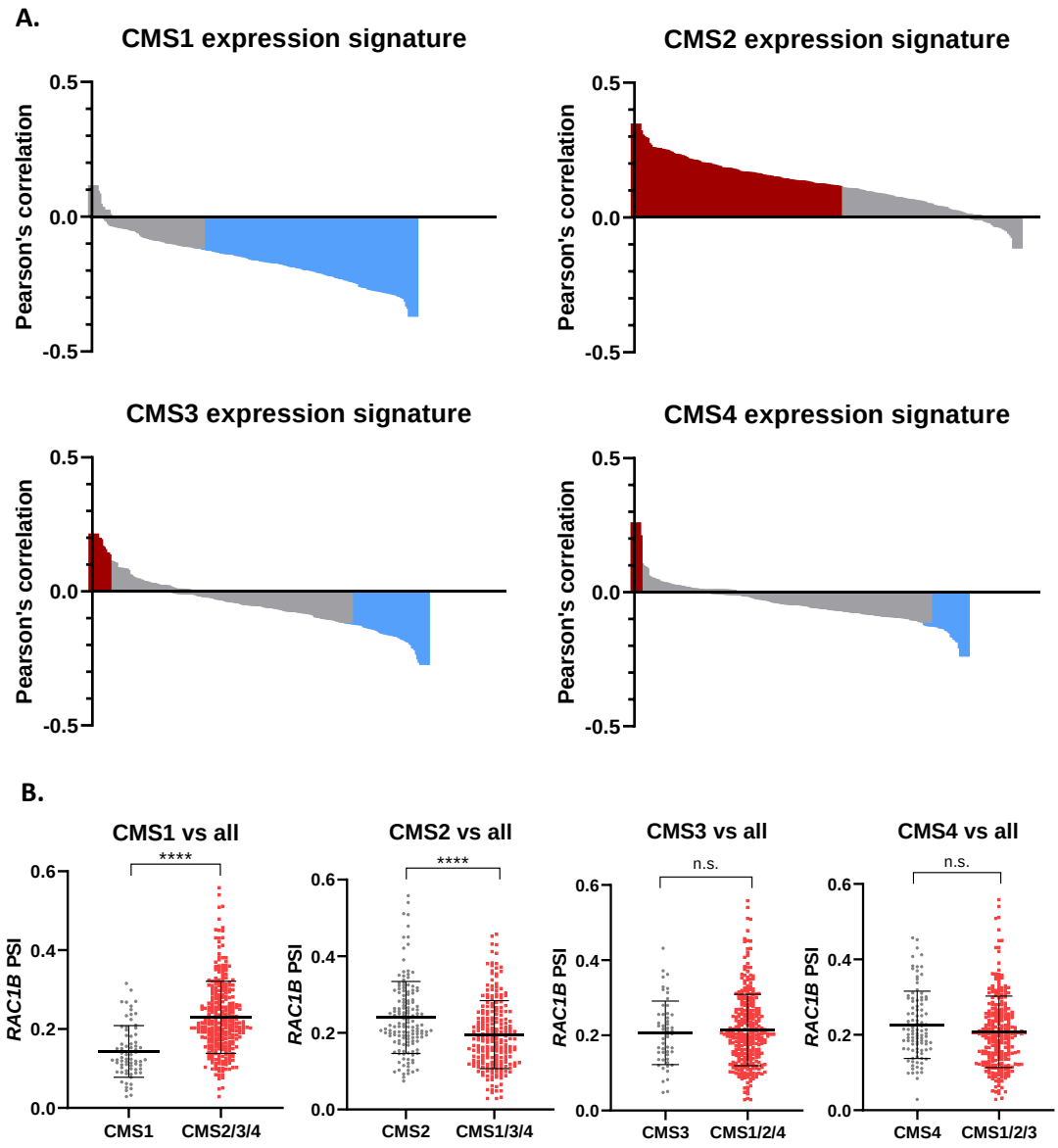
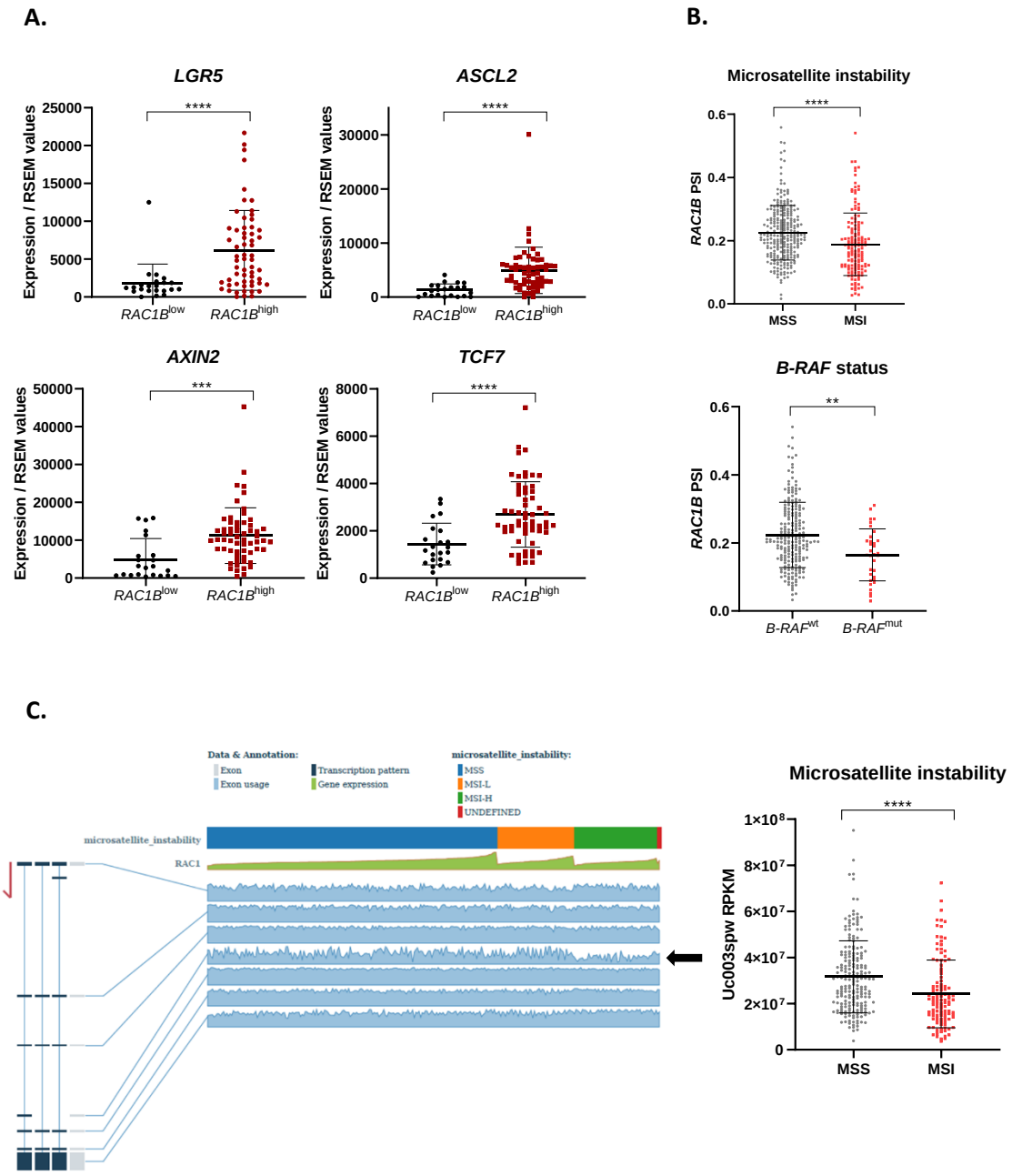


Figure 3.6: Expression of *RAC1B* correlates with CMS1 and CMS2 subtypes. A: Pearson's graphs correlating *RAC1B* PSI and genes defining CMS1, CMS2, CMS3 and CMS4 expression signature. Blue zone: significant negatively correlated genes, red zone: significant positively correlated genes, grey zone: not significant correlation. B: *RAC1B* PSI values classified comparing one CMS subtype vs the other subtypes.
Mann-Whitney test: ****= p≤0.0001, n.s. = not significant.



<u>Figure 3.7</u>: Featuring traits of CMS1 and CMS2 correlate with *RAC1B* expression. A: RSEM counts of the WNT/β-catenin target genes *LGR5*, *ASCL2*, *AXIN2* and *TCF7* comparing RAC1B^{low} and RAC1B^{high} groups. B: Microsatellite status and presence of *B-RAF* mutation plotted according to *RAC1B* PSI values. C: Microsatellite status of COAD TCGA tumours extracted from the TSVdb platform. Arrow marks information for *RAC1B* isoform, and this data is plotted in the adjacent graph.
Mann-Whitney test: **= p≤0.01, ***= p≤0.001, ****= p≤0.0001.

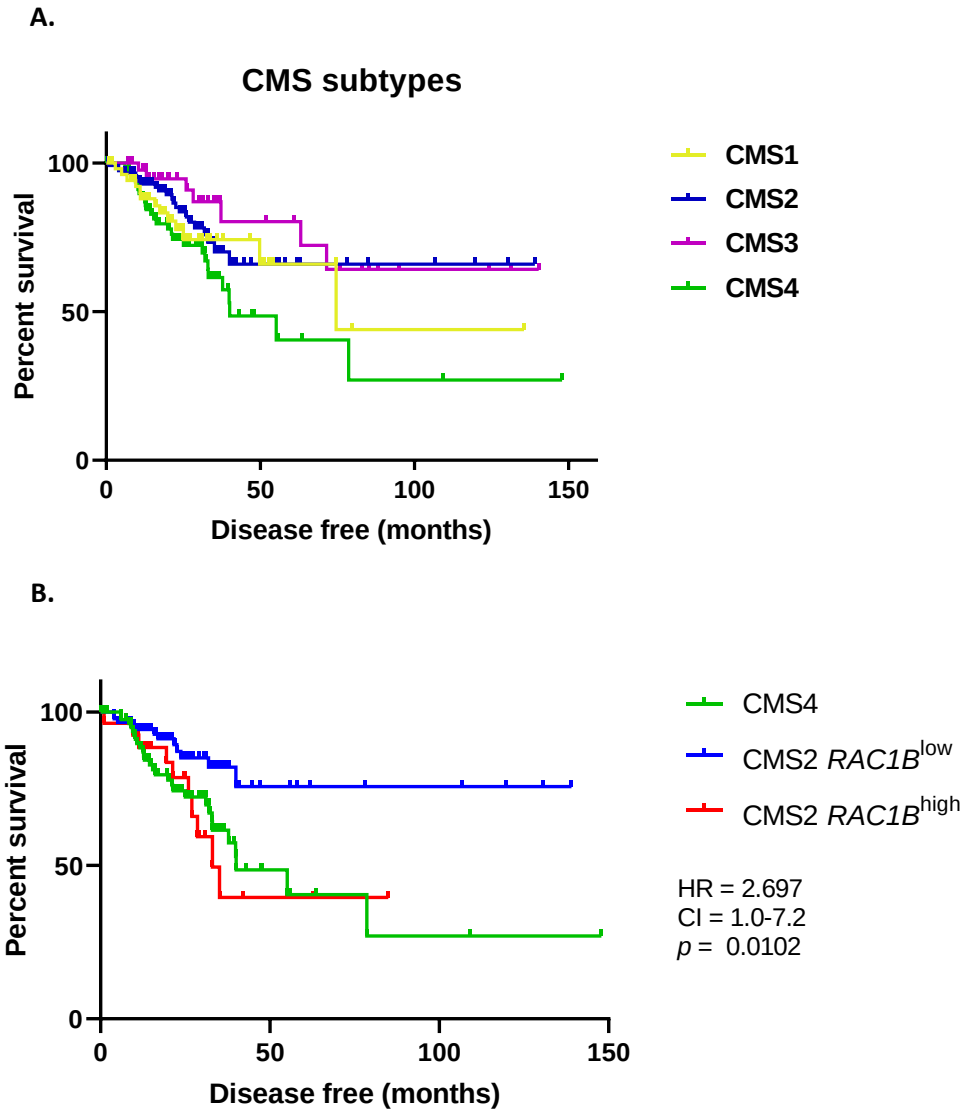


Figure 3.8: Expression of *RAC1B* subclassifies CMS2 survival outcome. A: Kaplan-Meier plot for disease-free survival of the CMS subtypes. B: CMS2 survival is plotted based on RAC1B^{low} and RAC1B^{high} tumours. CMS4 is illustrated as a comparative survival curve. Log-rank test comparing CMS2 RAC1B^{low} and CMS2 RAC1B^{high} survival curves. HR: Hazard Ratio, CI: Confidence Interval.

3.2.3 *RAC1B* PSI correlates with genetic alterations in cancer-driver genes

Analysing the *B-RAF* mutational status in relation to *RAC1B* overexpression paved the way to carry a similar search for other relevant oncogenes/tumour suppressor genes in CRC according to *RAC1B* PSI. Clinical data from TCGA only provides information about *B-RAF* and *K-RAS* mutations, but CBioPortal allows for the search of a manually entered list of genes within a data set (TCGA COAD) and informs about the kind of mutation it has taken place. A set of thirteen genes representative of the most commonly mutated cancer pathways were selected and *RAC1B* PSI was plotted relative to presence of absence of alteration in them (Figure 3.9B). Since the number of samples in this data set is larger, the reduced *RAC1B* expression previously observed in *B-RAF* mutated tumours is now more evident, with a stronger statistical power than before (Figure 3.9A). Other signalling pathways that when mutated exhibit lower *RAC1B* expression are the PI3K (by *PIK3CA* and *PTEN* genes) and the TGF-β signalling pathway (by *SMAD4* and *SMAD2*), which might support results evidenced in the following chapters. Notably, tumours where the *APC* gene is mutated express significantly higher levels of *RAC1B*, reinforcing Rac1b-related WNT signalling roles. *RAC1B* expression was also higher in tumours that carry *TP53* or *K-RAS* mutations but the functional relevance of such associations is unknown. Given the fact that these genes are amongst the most mutated genes in CRC, their mutational correlation with *RAC1B* support the rationale for further experiments to determine the relevance of these findings. Overall, the mutational analysis indicates *RAC1B* expression is highest in *APC* mutant tumours and lowest in tumours with *B-RAF* mutations, in agreement with the results of my subtyping analysis.

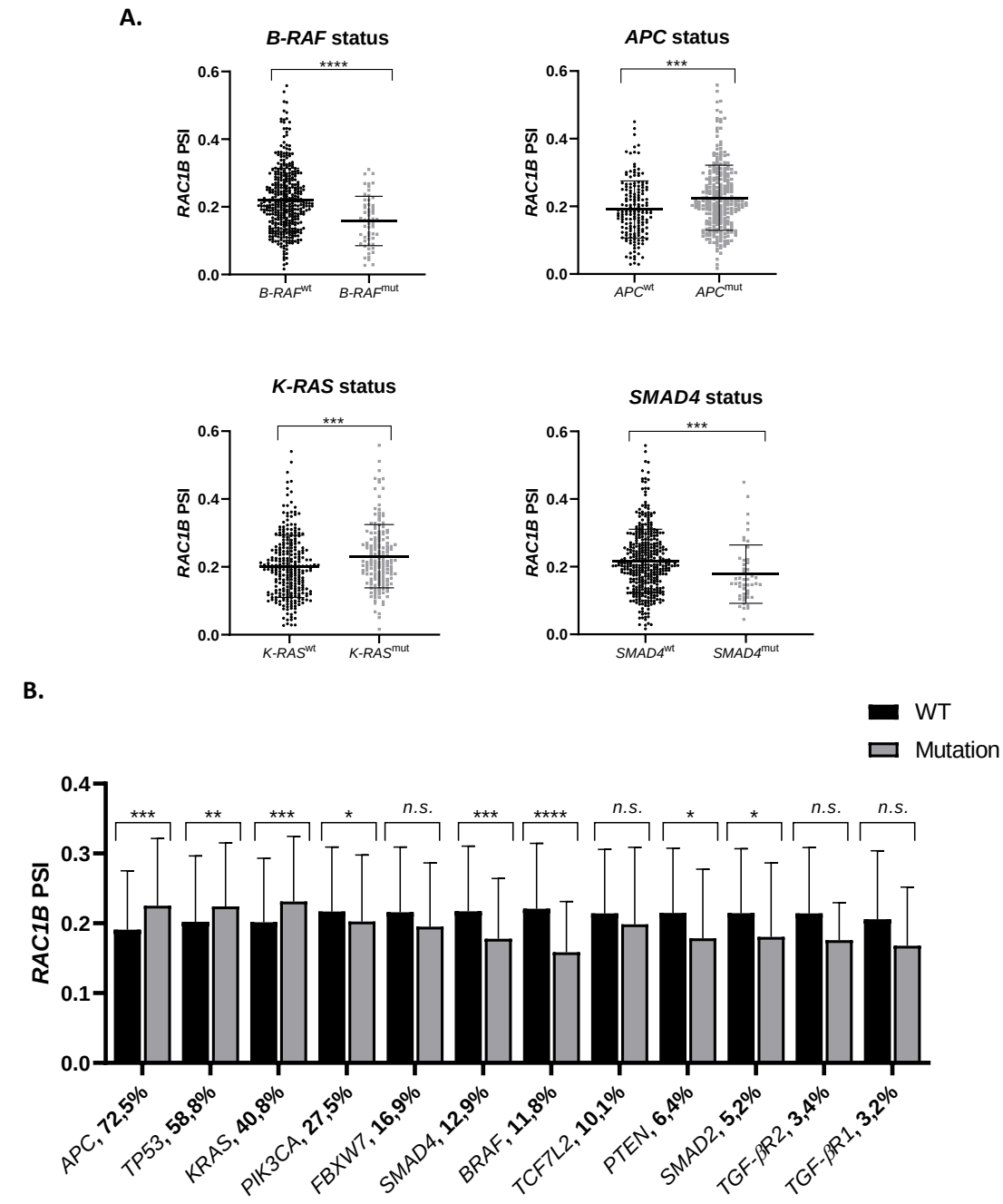


Figure 3.9: Mutational status of cancer-driver genes according to *RAC1B* expression. A: Individually plotted mutational data for *APC*, *SMAD4*, *B-RAF* and *K-RAS*. B: Array of twelve highly mutated genes in CRC, presented according their mutation status (WT or mutated) in relation to *RAC1B* PSI. Percentages indicate their incidence of mutation in CRC.
Mann-Whitney test: *= p≤0.05, **= p≤0.01, ***= p≤0.001, ****= p≤0.0001, n.s. = not significant, p>0.05.

3.2.4 RPPA identifies cellular processes associated with RAC1B expression

To conclude the *in-silico* study of Rac1b in cancer, I explored a different type of data also published by the TCGA. Reverse Phase Protein Array (RPPA) is a technique that allows efficient and high-throughput protein expression quantification by the binding of high-quality antibodies to a dot-blot membrane. Unlike RNA- or DNA-based assays, RPPA can detect pathway activation via post-translational modifications such as phosphorylation or protein cleavage. Based on the proteomic dataset provided by The Cancer Proteome Atlas (TCPA) web portal[454,455], level 4 data was downloaded from the TCGA COAD dataset. These data consisted of a list of protein names per each TCGA tumour, including phosphorylated or cleaved proteins, that were quantified using a "supercurve fitting" methodology, developed as well at the MD Anderson Cancer Centre. After overlaying TCGA codes with the codes from the *Rac1* PSI list, a Pearson's correlation test was run per all proteins. Out of 223, seventeen positive and twenty-one negative significantly correlated proteins resulted from the test, names of which are highlighted in Figure 3.10. There were some interesting targets, as the positive correlation with c-myc, SRSF1 (Rac1b's postulated positive regulator), INPP4B (phosphatase within the phosphatidylinositol pathway) and activation of YAP. On the contrary, within the negative correlated list there were a number of caspases (Caspase-3, -8 and cleaved-caspase-7) and cell check-point proteins (as activated Chk1) that would suggest a potential avoidance to cell-death programs. Finally, the most negative correlated protein is Prex1, a Rac1 guanine nucleotide exchange factor (GEF), which may indicate that high levels of Rac1b in tumours are not redundant with other mechanisms of Rac1 signalling activation. Therefore, analysing an additional TCGA dataset as RPPA has reinforced previous observations like the WNT signalling pathway correlation (c-myc) and downregulation of PI3K pathway, and suggested new processes to be evaluated like the Hippo-YAP signalling pathway or caspase-dependant cell death, both highly relevant during tumorigenesis.

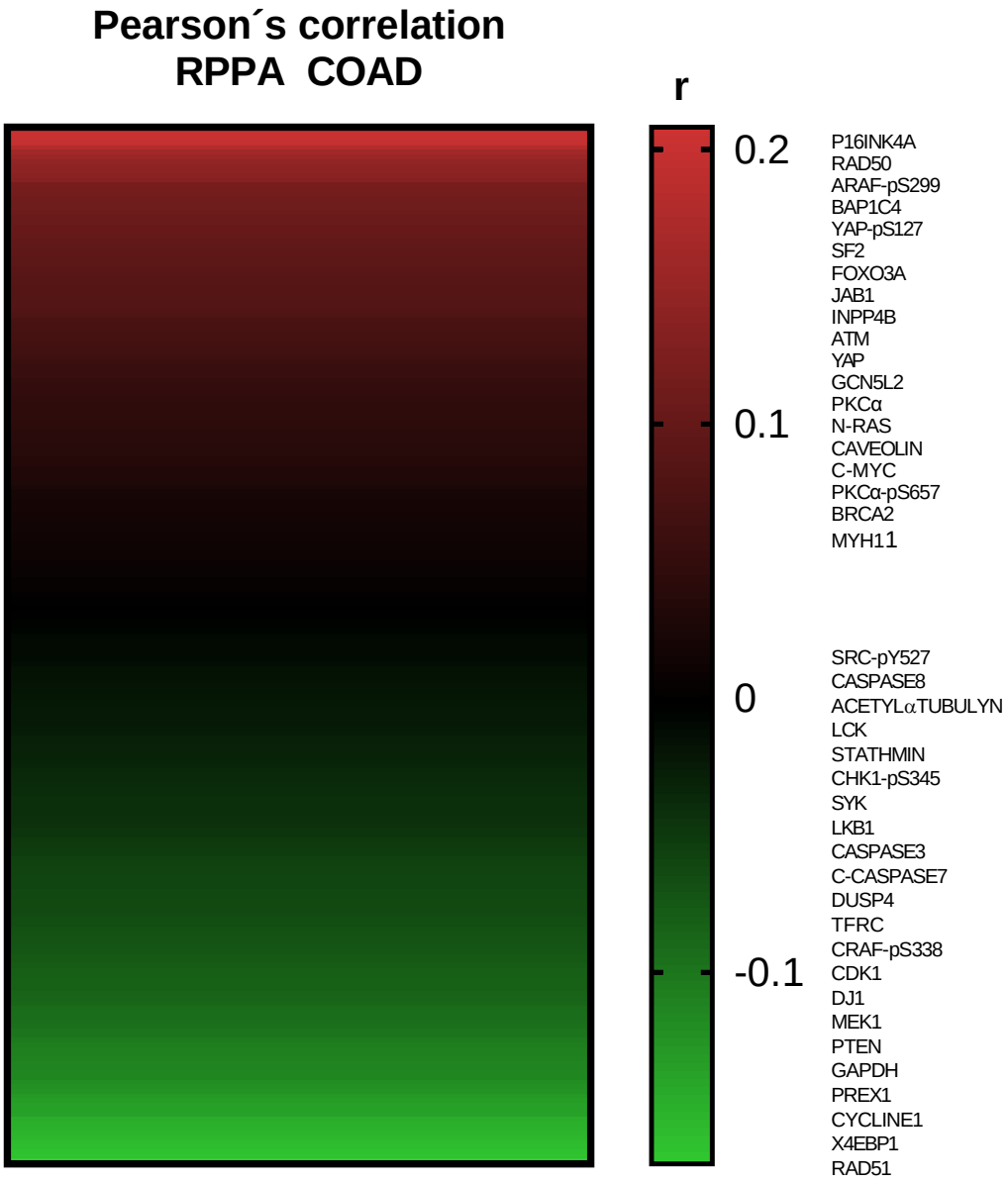


Figure 3.10: Heatmap correlation of TCGA RPPA and *RAC1B* PSI. Protein names refer to the most positive (red) and negative (green) correlated proteins with *RAC1B* expression.

3.2.5 Summary

The TCGA program has become of a great value to estimate the significance of a certain gene in cancer, to prognosticate the relationship with its expression and patient survival and to correlate with other genes or signalling pathways. As data sets were being compiled, many studies have been published gathering different cancer types together to analyse a cellular process[456,457], as well as to elucidate unknown cancer driver genes[458] or to predict the efficacy of a therapeutic agent[459]. Due to its ease of availability, it is the main dataset source for assessing human relevance of experimental findings and many projects have benefited from it. Utilising this dataset I find that *RAC1B* has an increased expression in tumours compared to normal tissue and that *RAC1* presented the antagonist result (Figure 3.4), highly suggesting Rac1b as a mechanism to maintain high Rac1 signalling in cancer. I then carried out a pan-cancer analysis by making use of the TCGA SpliceSeq web platform data. I showed that in seven out of twenty different cancer types, there was a significant increase of *RAC1B* expression in tumours compared to matching normal tissue (Figure 3.4D) and that COAD was among these seven. Dividing the COAD tumours according to its *RAC1B* expression (low or high) (Figure 3.5) showed a survival difference depending on *RAC1B* expression, since tumours within the high group have a worse overall survival and a disease-free survival compared to RAC1B^{low}. Accordingly, high *RAC1B* expression correlated with late stage disease (III and IV). These tumours were more aggressive given that they were more prone to invade the lymphatic system and blood and presented more metastasis compared to low *RAC1B* tumours. When trying to classify *RAC1B*$^{+ve}$ tumours to one of the fours CMS subtypes, I found a negative correlation with the CMS1 subtype and a positive correlation with the WNT high CMS2 subtype (Figure 3.6). Remarkably, the CMS2 is the one with the best survival prognosis but this is driven by the tumours with low *RAC1B* expression, given that CMS2- *RAC1B*^{high} presented a worse survival curve, similar to that of CMS4 (Figure 3.8). Mutational analysis of the top cancer-driver genes in CRC revealed some interesting correlations with *RAC1B* expression, such as the positive correlation with *APC*, *TP53* and *KRAS*, top three mutated genes in CRC, and the negative correlation with genes involved in the TGF-β and PI3K signalling pathways (Figure 3.9). Some of these were further supported by the RPPA analysis, which also suggested new cellular processes up to date not linked like cell death resistance or Hippo-YAP signalling (Figure 3.10). However, correlating RNA-base data (RNAseq) with Protein-based data (RPPA) might derive to some missassociation and further functional

validation is required before making any assumptions. Altogether, this study has reinforced the need to further investigate the function of Rac1b in cancer.

3.3 Studying Rac1b oncogenic function *in vivo*

Mouse models were the chosen method to study Rac1b cancer-related roles *in vivo*. In human tumours, I observed that *RAC1B* expression is increased in tumours that invade and metastasise. In order to validate that cancer murine models could emulate a similar situation, the expression of *Rac1b* was analysed in a set of CRC mice models, ranging from benign (homozygous deletion of *Apc,* APC) to the most invasive (*Apc* deletion along with homozygous deletion of *PTEN* and overexpression of *c-Myc,* APC PTEN CMYC). A preliminary analysis by qRT-PCR showed that intestine from mice that had lost both copies of *Apc* had significant more expression of *Rac1b* compared to *Apc* wild type mice (Figure 3.11). Moreover, this expression increased as mice acquired additional mutations like *TP53*, *PTEN* or a combination of *PTEN* and *c-Myc* mutations, which mimics a metastatic model. This suggests that upon oncogenic mutations, the murine intestine also increases *Rac1b* expression. Besides, Rac1b may be involved in driving the malignant phenotype of these models and the modulation of its expression could derive interesting yet unknown functions of Rac1b. Following the literature and my own TCGA research about Rac1b, I have postulated two different (and not mutually exclusive) hypotheses for its function in CRC. On one side, since its expression significantly increases immediately after *Apc* loss (Figure 3.3) and it seems to correlate with a WNT phenotype in tumours (Figure 3.7) I hypothesise that it facilitates *Apc* loss driven tumour initiation. On the other side, I have also observed that patients with tumours with high *RAC1B* expression have poor survival and are more prone to disease progression (Figure 3.5). Therefore, I hypothesise that Rac1b is also involved in driving tumour invasion. In order to answer whether Rac1b is either required or a promoter of these aforementioned processes, I used CRC animal models with the Villin Cre tamoxifen-inducible system to emulate different disease stages, modulate *Rac1b* expression and determine the phenotypic outcome. For *Rac1b* overexpression I utilised the previously described model developed by Zhou *et al.*, which consist of an inducible human Rac1b cDNA knocked into the Rosa26 locus (termed Rosa26-LSL-Rac1b). For loss of function experiments, knockout of the exon 3b was achieved by the insertion of loxP sites flanking the exon 3b ($Rac1b^{fl/fl}$). These

mice were generated by Douglas Strathdee and their phenotypic outcome has not been described yet. Both mice were crossed with VillinCre^{ERT2} mice, permitting tamoxifen inducible expression[408] (Figure 2.1).

3.3.1 Modelling early adenomas

Adenoma formation can be modelled by loss of one copy of the *Apc* gene. The model chosen was the *VilCreER^{T2}Apc$^{fl/+}$* (hereafter referred as APC)[427,447]. After tamoxifen induction, the Cre efficiently recombines along the intestine but *Apc* heterozygosis does not translate into an external phenotype. Stochastic loss of the remaining *Apc* copy allows tumour formation in the SI and LI. As of 100 days, mice start to present pale feet which is a sign of anaemia due to tumour burden. Some of them might present prolapses as a consequence of tumour formation in the distal colon and when the overall performance of the animal and anaemia have reached their endpoint threshold, mice are sacrificed with an average survival of 180 days. Tumours developed are mostly from the small intestine, ranging from 42 to 88 tumours and invasive tumours are rare[460]. To study the role of *Rac1b* gain and loss of function in tumour initiation, *Apc$^{fl/fl}$* mice were crossed with *VilCreER^{T2}Rosa26$^{lsl\text{-}Rac1b/lsl\text{-}Rac1b}$* (APC Rac1b^{OE}) and *VilCreER^{T2}Rac1b$^{fl/fl}$* (APC Rac1b^{KO}), respectively (Figure 2.1). Results from both cohorts correspond to chapter 4.

3.3.2 Modelling invasive adenocarcinoma

In order to generate an invasive CRC mice model, the APC model can be combined with other oncogenic mutations. A large number of models have been described[424,461,462] and for the present project, I will use a model that targets *TP53* for deletion along with heterozygous *Apc* deletion. Thanks to the O. Sansom lab, a *VilCreER^{T2}Apc$^{fl/+}$TP53$^{fl/fl}$* model is available. Its phenotype has not been described before but unpublished data from his lab observed a high frequency of invasive tumours and a reduction of survival compared to wild type *TP53* allele, which gives as a window to exacerbate or ameliorate the phenotype by *Rac1b* gain or loss of function.

Following the same methodology as with the APC cohort, *TP53* mutant mice were crossed with Rac1b^{OE} and Rac1b^{KO} mice (Figure 2.1). Using these models, my goal was to

determine whether the invasive tumour phenotype observed in these mice is altered by deletion or overexpression of *Rac1b*. Results from these cohorts are presented in detail in chapter 5. Depending on the outcome I will be able to determine whether Rac1b promotes or is required for tumour invasion and/or metastasis.

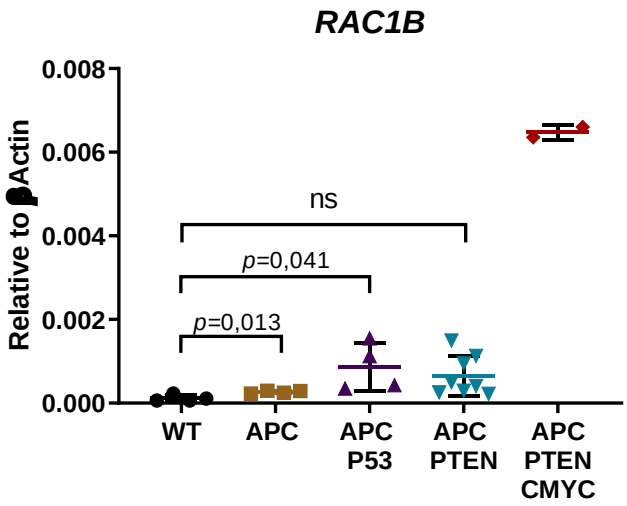


Figure 3.11: CRC mouse models emulate *Rac1b* expression pattern observed in human cancer. qRT-PCR for *Rac1b* in WT mice and mice with mutations in the following alleles: $APC^{fl/fl}$, $APC^{fl/fl}$ $TP53^{fl/fl}$, $APC^{fl/fl}$ $PTEN^{fl/fl}$, $APC^{fl/fl}$ $PTEN^{fl/fl}$ $c\text{-}Myc^{mut}$. T-test for significance: ns = not significant, p>0.05. N(WT, APC, APC P53)=4; N(APC PTEN)=8; N(APC PTEN CMYC)=2.

3.4 Discussion

To initially characterise Rac1b expression *in vivo* it was necessary to describe its intestinal localisation. An initial handicap found in the project was the inability of the Rac1b antibody to specifically recognise neither murine or human Rac1b protein. However, that problem was overcome by studying its expression at the RNA level. Both fractionation experiments and BaseScope staining demonstrated a preferential crypt location of *Rac1b* which significantly increased upon homozygous deletion of *Apc*. Unfortunately, I could not validate its subcellular location. Rac1b, as well as Rac1, have a PBR region in its sequence which harbour a nuclear localisation signal (NLS), pointing to a possible role in the nucleus. Nuclear function of Rac1 has extensively been reported, while only Bapat and colleagues observed Rac1b nuclear location, acting as a cofactor to translocate β-catenin and Dvl3[387]. Even though its membrane localisation is fairly accepted, further investigation within cellular compartments could render interesting insights. Nevertheless, these novel RNA-based studies provided and validated some previous hypotheses. On one side, its crypt-enriched position suggests that it might be involved somehow in maintaining proliferation at the crypt zone, either directly or by modulating crypt-enriched signalling pathways. On the other side, its overexpression upon *Apc* deletion validates the association of *Rac1b* expression with tumours. On top, it might also suggest a link with the WNT signalling pathway, given that Apc is its core negative regulator. Indeed, TCGA data analyses revealed a strong association between tumours expressing *RAC1B* and the CMS2, which is the canonical/WNT activated subgroup.

Analysis of TCGA data revealed a significantly poorer prognosis in patients with tumours expressing high levels of *RAC1B*, reflected in both disease-free survival and overall survival. Moreover, high levels of *RAC1B* is associated with a malignant phenotype. Intriguingly, this contradicts its high expression in CMS2 tumours, as CMS2 is one of the subtypes with best prognosis. However, *RAC1B* expression enabled a second subclassification within the CMS2: tumours with low *RAC1B* expression had good prognosis and tumours with high *RAC1B* expression had poor prognosis. Indeed, the latter had as poor survival as the CMS4. This indicates that, in spite of belonging to a canonical subgroup with a rather treatable cancer, Rac1b might be responsible for the dismal response of some of them. Consistent with this, a former CRC stratification by Sadanandam and colleagues identified a subgroup within a group analogue to CMS2 (the trans-amplifying subtype) which presented

a cetuximab-resistant phenotype (and EGFRi) and a poor prognosis[202]. Altogether, this might indicate that *RAC1B* expression can stratify CMS2 tumours according to their survival outcome and may function as a biomarker to anticipate treatment response.

Interestingly, another strong association with the CMS was the negative correlation with CMS1. To date, *RAC1B* expression was believed to associate with $BRAF^{V600E}$ mutated CRCs and accordingly, with tumours originated through the serrated pathway[407]. However, TCGA data revealed exactly the opposite, as it is also supported by its negative correlation with MSI status. Even though $BRAF^{V600E}$ mutation and *RAC1B* overexpression might cooperate to promote cell growth in an artificial setting, such as when *RAC1B* is ectopically overexpressed in CRC cell lines, TCGA data indicate that they do not associate in human tumours. Besides, given the significant correlation with *APC* mutations and *RAC1B* expression, *BRAF* negative correlation is somehow expected due to their mutational exclusivity.

The observation that *Rac1b* expression also increases in CRC mouse models as the model become more aggressive points to a potential approach for determining Rac1b function using a model that emulates what is seen in human tumours. Hence, my TCGA analyses have served as an initial basis to reinforce the Rac1b pro-tumorigenic role in CRC. It has identified two main functionally relevant traits: association with CRC progression and correlation with WNT/β-catenin-activated tumours. These findings are the basis for the functional experiments described in the following chapters.

Chapter 4: Rac1b function in adenoma tumour initiation

Mutations or loss in the *Apc* gene are considered a key initiating event in human CRC and a starting point for most mouse models. As a negative regulator, it is a core element within the WNT/β-catenin signalling pathway and overactivation of the pathway is a hallmark in CRC initiation. My previous TCGA analyses have shown a strong positive correlation between *Rac1b*, WNT-activated target genes (Figure 3.7) and *Apc* mutations (Figure 3.9). In addition, homozygous deletion of *Apc* in the murine intestine gives rise to an elevated expression of *Rac1b* (Figure 3.3), suggesting their inter-connection. Bearing all this in mind, the present chapter aims to describe the phenotypes observed following *Rac1b* gain and loss of function in an *Apc*-dependent mouse tumour model. Alongside, modulation of Rac1b expression will also be evaluated in normal tissue to study its potential role during regulation of homeostasis.

4.1 *Rac1b* expression is not required to maintain intestine homeostasis

In physiological conditions, normal intestine expresses low levels of Rac1b. I wanted to test whether this expression was necessary for normal functioning of the tissue and whether abrogation of Rac1b could cause any phenotypic effect. I used a mouse model where Rac1 exon 3b is flanked by loxP sites permitting specific deletion of *Rac1b* without altering *Rac1* expression. These mice were crossed with VilCreER^{T2} line generating cohort of *VilCreER^{T2}Rac1b$^{+/+}$* (WT) and *VilCreER^{T2}Rac1b$^{fl/fl}$* (RAC1B^{KO}) mice. Following tamoxifen induction and Cre expression, *lox-exon3b-lox* is excised and the remaining transcript only produces Rac1 protein (Figure 2.1). The effect of *Rac1b* deletion in the murine intestine was investigated by short term tamoxifen induction in WT and RAC1B^{KO} mice. As described in the previous section, the LI expresses higher levels of *Rac1b* than the SI. Consequently, *Rac1b* knockout in the SI is hardly distinguishable by qRT-PCR, while the colon presents a significant deletion of *Rac1b* with about 3-fold decrease (Figure 4.1A). Although I cannot corroborate the knockout in the SI by qRT-PCR, I described in the previous section how expression of *Rac1b* is virtually undetectable after acute tamoxifen induction by means of BaseScope (Figure 3.3), validating the Cre knockout system in both SI and LI. I then focused in the SI to study the stem cell population and the WNT signalling pathway by qRT-PCR and Paneth and

goblet cells by IHC in order to evaluate homeostatic changes made by *Rac1b* deletion. Figures 4.1B and C illustrate a characterization of all cell populations and WNT activity which is nearly equal in both groups, and consequently, there is no change in proliferation upon *Rac1b* deletion. Thence, abolition of *Rac1b* expression in normal intestine does not cause a phenotypic homeostatic alteration.

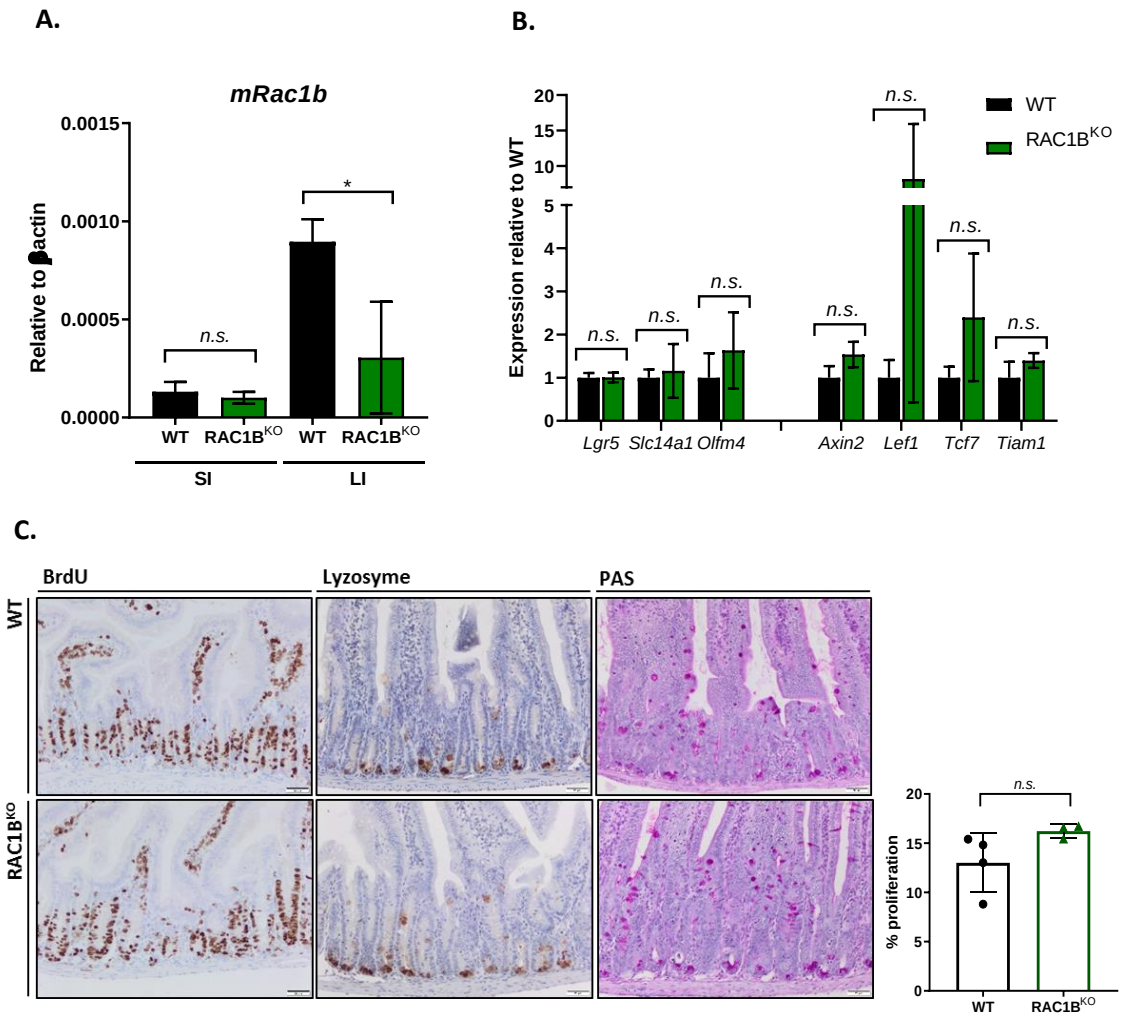


Figure 4.1: Deletion of *Rac1b* in the normal murine intestine. A: Validation of *Rac1b* depletion by qRT-PCR from SI and LI tissue. B: Analysis of stem cell markers (*Lgr5*, *Slc14a1* and *Olfm4*) and WNT/β-catenin target genes (*Axin2*, *Lef1*, *Tcf7* and *Tiam1*) RNA expression in SI tissue. C: Evaluation of cell proliferation (Brdu), Paneth (Lyzosyme) and goblet cells (PAS) by staining of the SI. Adjacent graph presents the proliferation score for at least 50 half crypts in each mouse. T-test: *= p≤0.05, n.s. = not significant, p>0.05. N= 4 vs 4.

4.2 Deletion of *Rac1b* extends mice survival through two potential mechanisms

4.2.1 Ageing cohort with *Rac1b*-deficient mice

To our knowledge, there is no *in vivo* evidence to demonstrate whether Rac1b could delimit or be required to drive cell transformation and tumour formation by loss of function. Hence, we decided to test *in vivo*, for the first time, how deletion of *Rac1b* but not *Rac1* could affect early stages of tumour development. Taking *Apc* as the tumour initiating event, *VilCreER^{T2}Apc$^{fl/+}$* mice were crossed to the Rac1b floxed miced to generate cohorts of *VilCreER^{T2} Apc$^{fl/+}$Rac1b$^{+/+}$* (APC) or *VilCreER^{T2} Apc$^{fl/+}$Rac1b$^{fl/fl}$* (APC RAC1B^{KO}) mice (Figure 2.1). Once they were over 8 weeks old or reached 20 grams, mice were injected with one dose of 80mg/kg of tamoxifen and aged until they became symptomatic of disease. In collaboration with Caroline Billard, mice were closely monitored twice a week and changes in their weight or their performance were annotated. Pale feet, a sign of anaemia due to tumour burden, were noticeable after 100 days of induction and termination was required approximately two weeks after. As shown in figure 4.2A, APC RAC1B^{KO} mice presented with significantly increased survival compared to controls, with a median survival of 170 days compared to 138 respectively (p=0.0395). The prolonged survival of APC RAC1B^{KO} mice is due to a reduction in tumorigenesis, since they developed significantly fewer tumours than the APC group (41 vs 25 tumours for APC and APC RAC1B^{KO} respectively, p=0.0058, figure 4.2C). Importantly, the lack of differences in tumour burden states that all mice were terminated at the same clinical endpoint criteria (Figure 4.2D). In this model, tumorigenesis is predominantly localised to the SI; the reduction in tumour number occurred in the most proximal regions of the intestine (duodenum and jejunum) (p=0,00023 and p=0.0026 respectively, Figure 4.2E). Therefore, this data demonstrates that deletion of *Rac1b* in the murine intestine significantly reduces tumour formation and prolongs survival.

I have previously confirmed that the model efficiently deletes *Rac1b* expression without altering *Rac1* (Figure 4.1). However, I did not test a long-term deletion in an *Apc* deficient context. In order to validate that suppression of *Rac1b* is maintained in the tumours and to establish the relationship between decreased tumorigenesis and *Rac1b* knockout, RNA from SI normal tissue and tumour from APC and APC RAC1B^{KO} mice was extracted and analysed by qRT-PCR. As expected, *Rac1* expression is unaltered in the APC RAC1B^{KO} group and it also remains constant after losing *Apc* (Figure 4.2F). Of note, it does slightly but

significantly increase in the tumours without *Rac1b*, suggesting a potential compensatory effect (average 0.037 vs 0.042 $\log 2^{\Delta CT}$ respectively, p=0.013). *Rac1b* expression is increased 4.5-fold in APC tumours compared to normal tissue (p=0.0044, Figure 4.2G). It is successfully deleted in the normal SI of APC $RAC1B^{KO}$ mice compared to SI APC (p=0.0093) but interestingly, there are two outcomes in tumours from the APC $RAC1B^{KO}$ mice. While 6 out of the 11 tumours analysed have efficiently deleted *Rac1b*, 5 of them still express *Rac1b* at comparable levels to APC mice. This might suggest a positive selective pressure maintaining Rac1b expression for efficient tumorigenesis.

Therefore, these results validate the previous observations with short-term experiments: *Rac1b* expression is increased upon WNT signalling activation. They also demonstrate for the first time a functional requirement for Rac1b during the formation of intestinal tumours.

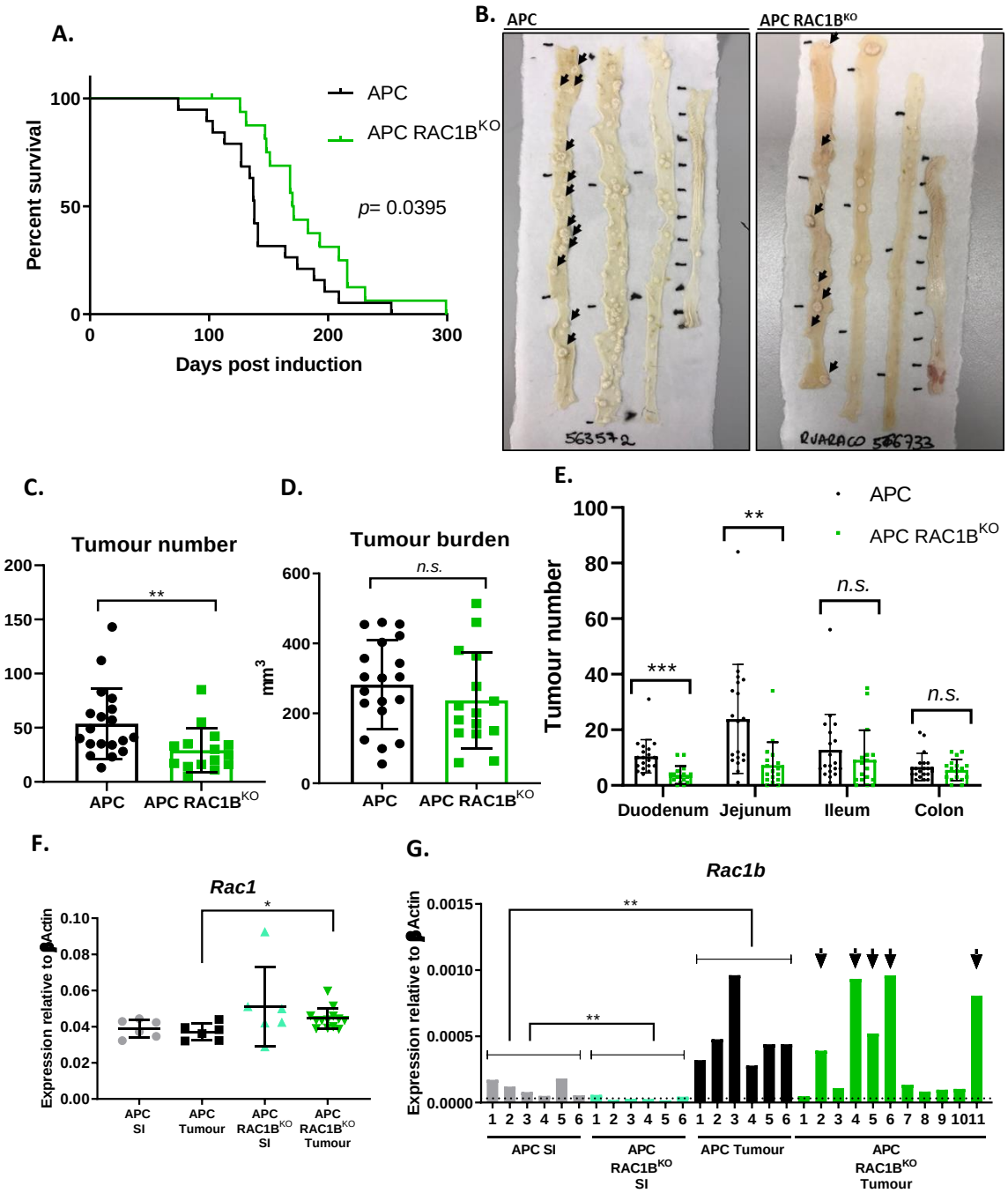
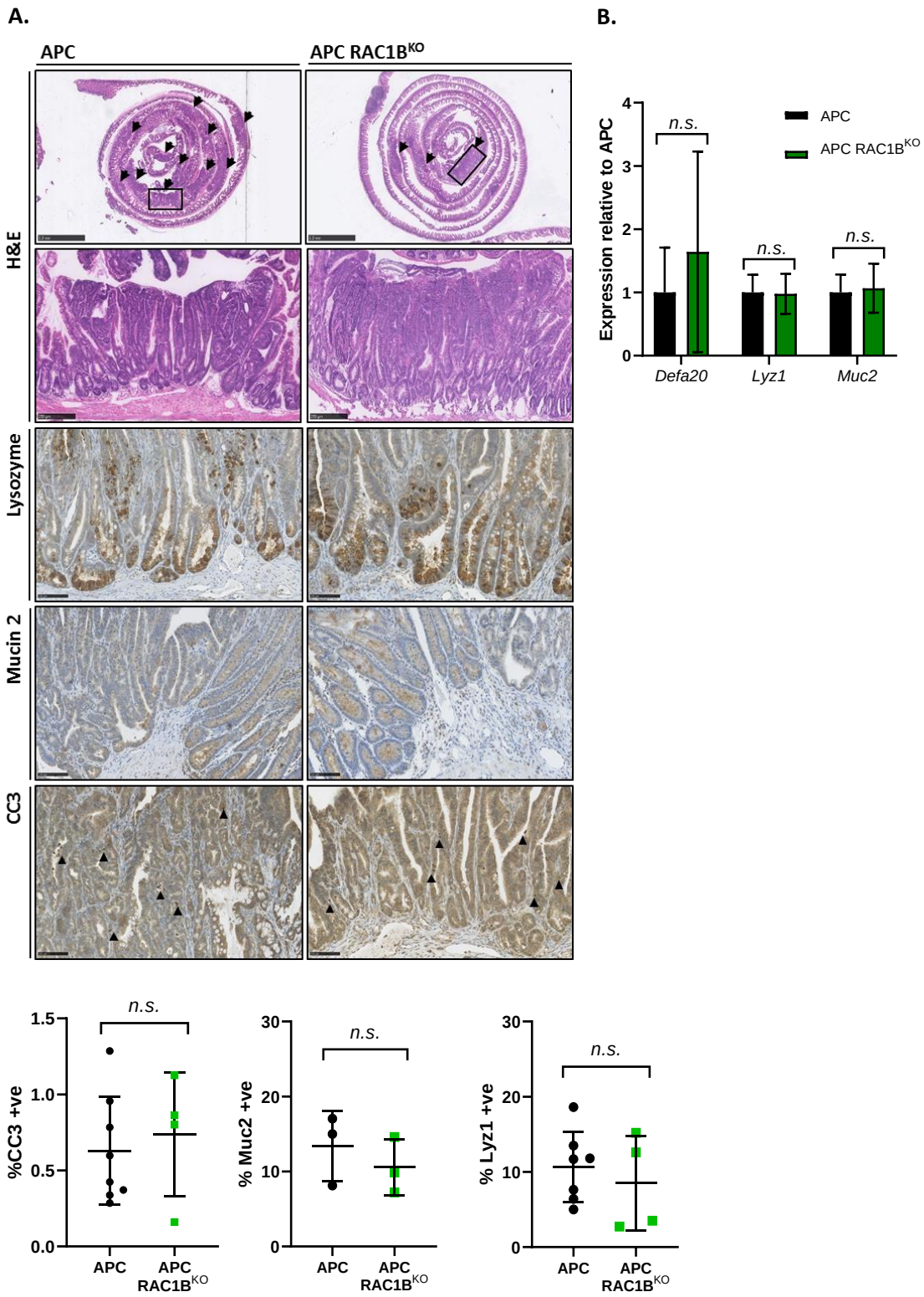


Figure 4.2: Deletion of *Rac1b* improves mouse survival and reduces intestinal tumorigenesis. A: Kaplan-Meier survival plot for APC and APC RAC1B^{KO} cohort mice (Log-rank test). B: Representative pictures of APC and APC RAC1B^{KO} intestines fixed on methacarn during dissection. Arrows mark developed tumours on the first 15 cm of the SI. C, D and E presents quantification of intestinal tumorigenesis, scored during dissection. C: Number of tumours developed per mouse (Mann-Whitney test, **= p≤0.001). D: Sum of diameters of developed tumours in each mouse. E: Tumour number presented according to the region where they originated. F: Expression of *Rac1* in both SI tumour and matched normal tissue for each genotype by qRT-PCR. G: RNA expression of *Rac1b* presented as individual values in both SI tumour and matched normal tissue. Arrows highlight samples that maintained *Rac1b* expression. Mice in the cohort (n): APC = 19, APC RAC1B^{KO} = 15. T-test: *= p≤0.05, **= p≤0.01, ***= p≤0.001, n.s. = not significant, p>0.05. N(qRT-PCR) 6 vs 6. Cohort conducted alongside Caroline Billard.

4.2.2 Histological study of the tumour phenotype

To understand mechanistically how *Rac1b* abolition impaired tumorigenesis and favoured mice survival, a histological characterization of the tumours was conducted. Immunohistochemical (IHC) analysis using markers for Goblet cells (Muc2) and Paneth cells (Lyz1) showed no differences in these cell populations, as neither did their corresponding qRT-PCR (Figure 4.3A and B). Besides, nuclear β-catenin IHC was used to validate high WNT activation within the tumours, which showed no differences between genotypes (Figure 4.3C). We then interrogated whether there were differences in tumoral cell death or proliferation. Apoptosis was analysed using cleaved-caspase 3 IHC and both APC and APC RAC1B^{KO} tumours had similar apoptotic rates (Figure 4.3A). However, when cell proliferation was scored by BrdU incorporation, *Rac1b*-deleted tumours had a significant reduction in BrdU^{+} cells compared to APC tumours (33.55% and 24.39% of proliferation for APC and APC RAC1B^{KO} respectively, p=0.0031, Figure 4.3D). Notably, some of the tumours scored within the APC RAC1B^{KO} group still express *Rac1b*, indicating that the reduced proliferation in these mice could be more dramatic if only *Rac1b*-deficient tumours were scored. Therefore, deletion of *Rac1b* in *Apc*-deficient tumours impairs cell proliferation, which likely translates into decreased tumorigenesis.


A.
B.
APC
APC RAC1B^KO
H&E
Lysozyme
Mucin 2
CC3
Expression relative to APC
n.s.
Defa20
Lyz1
Muc2
%CC3 +ve
% Muc2 +ve
% Lyz1 +ve
APC
APC RAC1B^KO

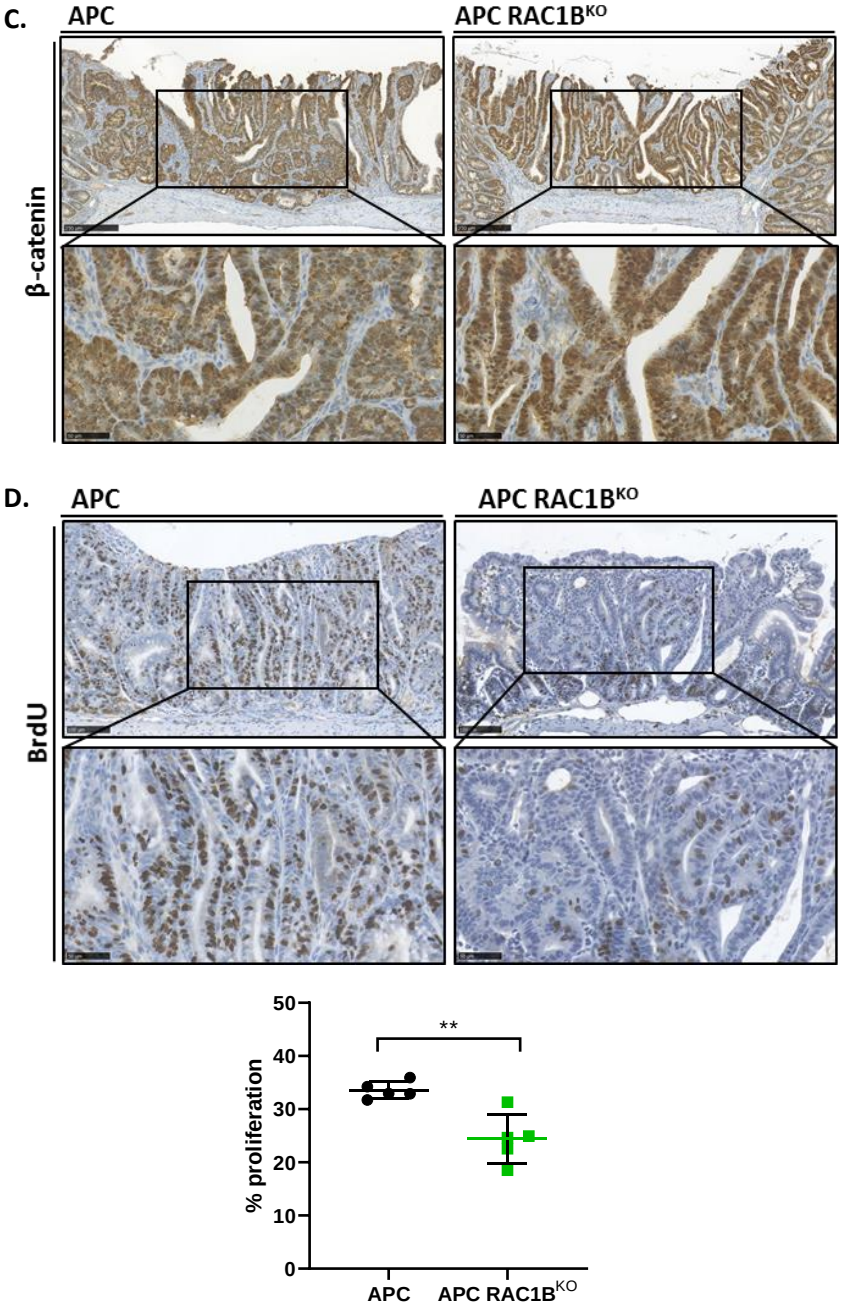


Figure 4.3: ***Rac1b*** **deletion reduces tumour proliferation.** A: Histological analyses of cohort tumours. First images offer an overview of SI Swiss-roll. Tumours in the section are highlighted with arrows and tumours within the rectangle are amplified underneath, both H&E stained. Images below show staining of Paneth cells (Lysozyme), goblet cells (Mucin 2) and cleaved caspase 3 (CC3). Average percentage of positive cells within a tumour is plotted underneath. Scores conducted with QuPath. B: qRT-PCR for Paneth (*Defa20* and *Lyz1*) and goblet cells (*Muc2*) from tumour RNA. C: Nuclear β-catenin IHC of tumours at 10X and 40X. D: BrdU staining at 10X and 20X. Proliferation scoring was conducted with QuPath and it is plotted in the graph. Analyses performed alongside Caroline Billard and Kevin Myant.
T-test: **= p≤0.01, n.s. = not significant, p>0.05. N(qRT-PCR) = 3 vs 3; N(CC3, Lyz) = 7 vs 4; N(Muc2) = 3 vs 3; N(βCat, BrdU) = 5 vs 5.

4.2.3 The WNT signalling pathway is modulated by *Rac1b*

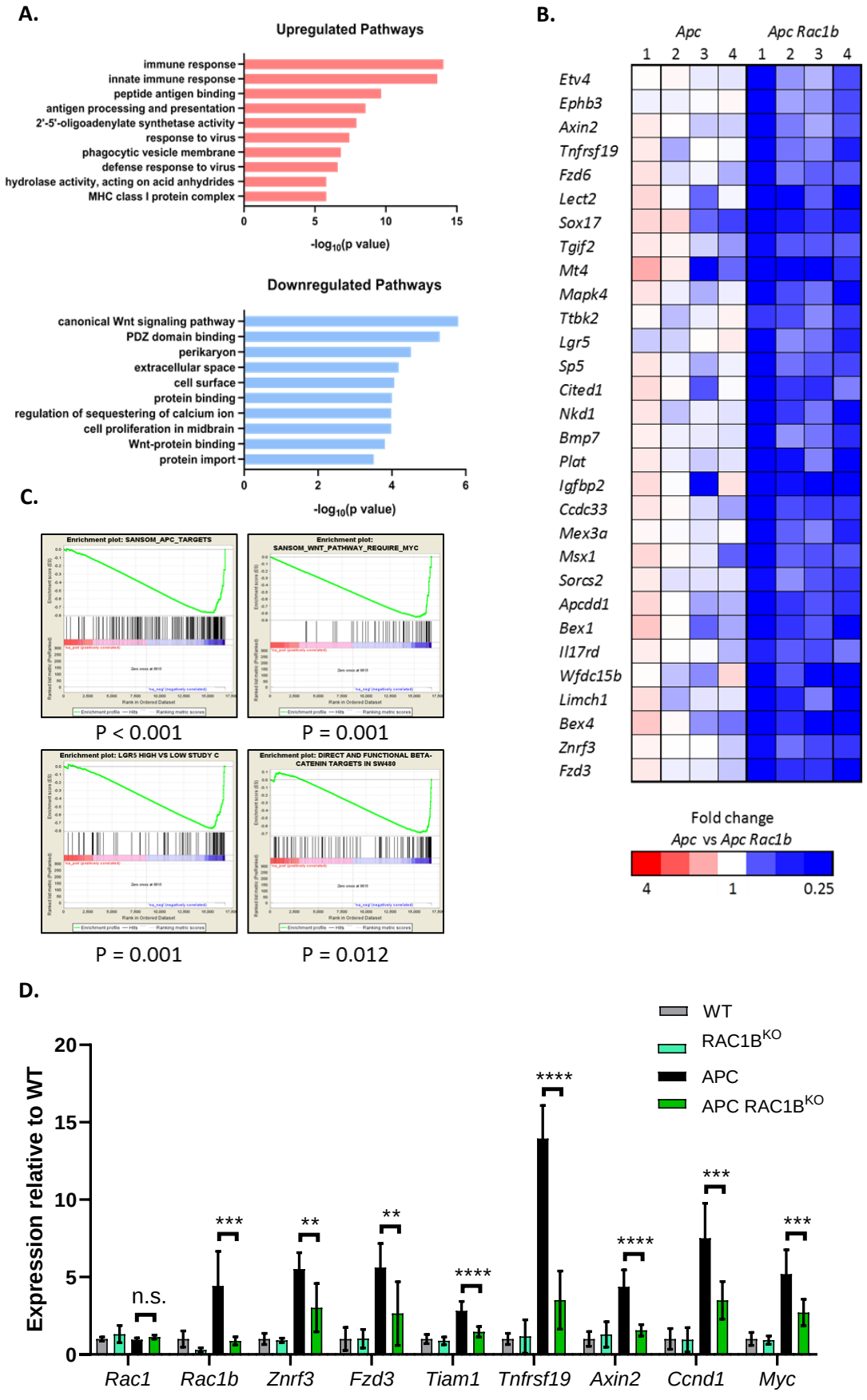
4.2.3.1 RNAseq analysis

Many signalling pathways could be implicated in the decreased cell proliferation observed. In order to explore from an unbiased perspective the altered genes following *Rac1b* downregulation, an RNA-seq comparing 4 APC and 4 APC RAC1B^{KO} tumours (with validated *Rac1b* deletion) was conducted. 558 differentially expressed mRNAs were identified (fold change>1.5, q value<0.05), of which 294 were significantly overexpressed and 264 were significantly downregulated (list of genes and FastQC details on Appendix 2.1 and 2.2, respectively). Gene ontology enrichment analysis of the identified data set demonstrated that immune regulatory processes such as immune and innate immune response, peptide antigen binding or antigen processing and presentation were among the top overrepresented pathways (Figure 4.4A). Interestingly, regulation of the canonical WNT signalling pathway was the most downregulated pathway in *Rac1b*-deficient mice, along with cell surface, protein import, cell proliferation in the midbrain and WNT-protein binding (Figure 4.4A and B). Further gene set enrichment analysis (GSEA) querying for WNT/β-catenin and APC target genes as well as for Lgr5^{+}/stem cells expression profile revealed a strong significant negative association with genes downregulated in our RNAseq data set (Figure 4.4C) (Appendix 3). For validation of our data set, qRT-PCR for WNT-related target genes was carried out on tumour samples from APC and APC RAC1B^{KO} mice, as well as on normal small intestine tissue from wild type mice (WT and RAC1B^{KO}). As it is shown in figure 4.4D, deletion of *Rac1b* in tumour samples let to a significant downregulation in our queried genes, while expression in matched normal tissue was not different between genotypes. Altogether, this points to the attenuation of WNT signalling activity as a possible mediator through which *Rac1b* deletion reduces tumour proliferation.

4.2.3.2 Acute tamoxifen induction reaffirms WNT-driven phenotype

To confirm the requirement of Rac1b for efficient WNT signalling following *Apc* loss, I analysed target gene expression in the SI of mice following acute *Apc* deletion. qRT-PCR from the SI of *VilCreER^{T2} Apc$^{fl/fl}$* (APC$^{-/-}$) and *VilCreER^{T2} Apc$^{fl/fl}$ Rac1b$^{fl/fl}$* (APC$^{-/-}$ RAC1B^{KO}) mice five days post induction showed a significant downregulation of a subset of WNT-targeted genes

such as *Tiam1*, *Axin2* or *Lef1* (Figure 4.4E), indicating that WNT signalling pathway activation immediately following *Apc* loss is mediated by *Rac1b* expression. Moreover, epithelial intestinal cells from these mice were extracted and cultured as 3D organoids for conducting a clonogenicity assay. Briefly, epithelial extraction protocol (material and methods chapter) was conducted and crypt-enriched cells from last fraction were digested with Accutase® and embedded as single cells in Matrigel®. The number of organoids formed was scored after four days of plating. As is shown in figure 4.4F, $APC^{-/-}$ $RAC1B^{KO}$ organoids have a worse clonogenic capacity compared to *Rac1b*-expressing organoids (p=0.028), suggesting that WNT downregulation translates into a reduced stem cell function and tumour initiating capacity.


A.
Upregulated Pathways
immune response
innate immune response
peptide antigen binding
antigen processing and presentation
2'-5'-oligoadenylate synthetase activity
response to virus
phagocytic vesicle membrane
defense response to virus
hydrolase activity, acting on acid anhydrides
MHC class I protein complex
0
5
10
15
-log10(p value)
Downregulated Pathways
canonical Wnt signaling pathway
PDZ domain binding
perikaryon
extracellular space
cell surface
protein binding
regulation of sequestering of calcium ion
cell proliferation in midbrain
Wnt-protein binding
protein import
0
2
4
6
-log10(p value)
B.
Apc
Apc Rac1b
1 2 3 4 1 2 3 4
Etv4
Ephb3
Axin2
Tnfrsf19
Fzd6
Lect2
Sox17
Tgif2
Mt4
Mapk4
Ttbk2
Lgr5
Sp5
Cited1
Nkd1
Bmp7
Plat
Igfbp2
Ccdc33
Mex3a
Msx1
Sorcs2
Apcdd1
Bex1
Il17rd
Wfdc15b
Limch1
Bex4
Znrf3
Fzd3
Fold change
Apc vs Apc Rac1b
4
1
0.25
C.
Enrichment plot: SANSOM_APC_TARGETS
P < 0.001
Enrichment plot: SANSOM_WNT_PATHWAY_REQUIRE_MYC
P = 0.001
Enrichment plot: LGR5 HIGH VS LOW STUDY C
P = 0.001
Enrichment plot: DIRECT AND FUNCTIONAL BETA-CATENIN TARGETS IN SW480
P = 0.012
D.
WT
RAC1B^KO
APC
APC RAC1B^KO
Expression relative to WT
0
5
10
15
20
n.s.
***
**
**
****
****
***
***
Rac1
Rac1b
Znrf3
Fzd3
Tiam1
Tnfrsf19
Axin2
Ccnd1
Myc

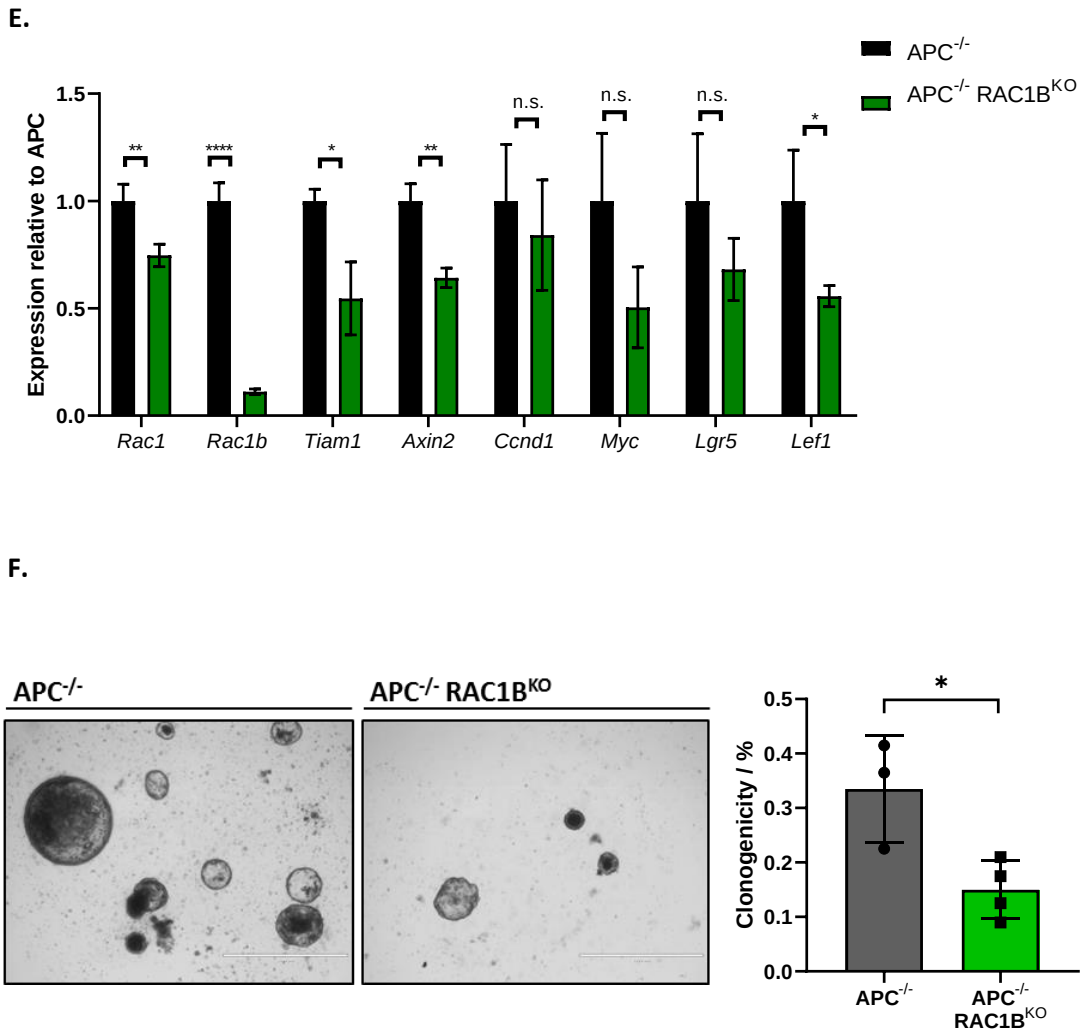


Figure 4.4: Identification of a WNT signature upon *Rac1b* deletion. Differentially expressed genes between APC and APC RAC1B^{KO} tumours identified by RNAseq were analysed through InnateDB database. Figure A presents the top ten pathways upregulated and downregulated in APC RAC1B^{KO}. B: Heatmap to illustrate the expression of WNT-related RNAseq target genes among tumour replicates. C: Gene set enrichment analysis (GSEA) of the RNAseq data set, utilising four different WNT-related gene sets. Genes are ranked based on their differential expression between the two cohorts. Black bars represent the ranked gene position, indicating a strong downregulation in all gene sets analysed. D: Validation of the WNT signature by qRT-PCR comparing tumours and matching normal tissue, the latter showing no statistical differences. E: qRT-PCR for WNT target genes on short-term APC$^{-/-}$ induced mice, presenting the same phenotype as endpoint tumours. F: Clonogenicity assay of APC$^{-/-}$ and APC$^{-/-}$ RAC1B^{KO} organoids. Number of clones formed after 4 days is plotted in the graph.
T-test: *= p≤0.05, **= p≤0.01, ***= p≤0.001, ****= p≤0.0001, n.s. = not significant, p>0.05.
Experiments performed in collaboration with Kevin Myant and Patrizia Cammareri. N = 4 vs 4; N(organoids) = 3 vs 3.

4.2.4 BioID protein interaction study reveals unknown Rac1b-interacting partners

A large number of genes involved in the WNT signalling pathway are downregulated in tumours lacking *Rac1b* and Gene Set Enrichment Analysis (GSEA) using WNT- and Apc-related gene lists demonstrates a strong negative correlation between this pathway and *Rac1b*-deficient tumours. Taking these data together with the results from the TCGA analysis (which revealed a positive correlation with Rac1b and CMS2 tumour subtype, the canonical WNT-activated subtype), there is a strong indication that Rac1b might enhance WNT activation following *Apc* loss. However, it is unknown how this regulation could be mediated. In order to gain insights into Rac1b downstream targets and its implication in other signalling pathways, a protein-protein interaction (PPI) experiment was conducted using BioID.

4.2.4.1 BioID experiment layout

BioID is a proximity-dependent biotin identification approach to uncover interacting partners of a given protein[463]. It relies on the fusion of a promiscuous mutated enzyme called BirA*, to a bait protein or protein of interest. Originally, BirA is a prokaryotic biotin ligase isolated from *Escherichia coli*. Due to its mutation, BirA* will biotinylate without specificity any interacting and vicinal proteins within a labelling radius of ~20nm upon exogenous biotin addition[464]. After biotin exposure, biotinylated proteins are pulled-down via streptavidin affinity and identified by Mass Spectrometry (MS) (Figure 4.5A). Compared to conventional methods to study PPIs such as affinity purification, yeast two-hybrid (Y2H) or fluorescence resonance emission transfer (FRET), BioID is an advantageous system to detect weak and transient interactions and study PPIs on their natural cellular environment avoiding solubility problems.

I made use of the BioID2 system to investigate Rac1b interacting partners, aiming to outline the mechanism through which it regulates the WNT signalling pathway and other as yet unknown downstream pathways. In addition, Rac1-containing BioID vectors were also generated in order to compare both interactomes and to establish specific Rac1b targets.

4.2.4.2 Rac1b and Rac1 cloning strategy

I used a BioID2 plasmid from Addgene® with Myc as a cellular tag in the N-terminus and cloning sites in the C-terminus for our protein (myc-BioID2-bait). Regarding the proteins of interest, *Rac1b* was purchased as a mouse cDNA from GeneScript® and amplified by PCR. In contrast, given that *Rac1* is highly and ubiquitously expressed in all tissues, its sequence was extracted by PCR from a piece of SI normal tissue. Since I aimed to eventually use this technique in organoids, which would only allow lentivirus transfection, I cloned the fused myc-BioID-Rac1b and myc-BioID-Rac1 vectors into a lentiviral vector (pLJM1-EGFP from Addgene®) to ultimately generate viruses (Figure 2.3). Cloning of a free-bait myc-BioID vector into the lentiviral vector was used as a BirA* control and an empty lentiviral vector (only pLJM1-EGFP) was used as a transfection control.

4.2.4.3 Optimization of BioID technique in CMT93 cell line

Once cloning was successful and sequences validated, pLJM1 vector with myc-BioID2-Rac1b (from now on, BirA-Rac1b), myc-BioID2-Rac1 (hereafter BirA-Rac1) and the control myc-BioID2 (BirA) were transiently transfected into mouse rectal carcinoma cells, the CMT93 cell line, by Lipofectamine® 2000 transfection reagent to verify correct protein expression and BirA biotinylation ability. Following Roux *et al.* protocol[465], 50µM of Biotin was added to the cell media when cells reached their transfection peak after about 40 hours. Although they recommended 24h of biotin treatment, I wanted to see whether biotinylation could be detected after shorter periods of time, which might then diminish potential contaminating protein background during MS. Cells were collected for RNA and proteins were extracted after 6 and 19 hours of treatment and both qRT-PCR and Western blot (WB) were conducted. As shown in figure 4.5B, there was a robust overexpression of mouse *Rac1* and *Rac1b* in Bir-Rac1 and Bir-Rac1b vectors respectively. Additionally, Western blots were used to evaluate protein biotinylation by Streptavidin-HRP antibody and protein bait overexpression by anti-myc-tag (9B11) antibody. As shown in figure 4.5D, biotinylation is already detectable after 6 hours only in the BirA-containing vectors and streptavidin bands were increased and more defined after 19 hours of biotin treatment. As expected, BirA has a higher streptavidin signal due to its unspecific protein affinity. The size of the myc-tag band will depend on the total size of the fusion protein. For the control vector (BirA) will be the sum of the myc-tag (about

8kDa) plus the BioID2 (23kDa), so a band at about 31kDa will be expected. However, for Rac1 will be 50kDa (19kDa Rac1) and about 53kDa for Rac1b. Figure 4.5C presents the WB films showing the expected size bands, indicating that protein overexpression and translation has been correct. The weak upper band in the BirA lane might probably be unspecific binding due to high antibody concentration, since subsequent WBs using a more diluted antibody concentration do not present this double band (data not shown). Therefore, both vectors and control were able to express their fusion protein and biotinylate in the CMT93 cell line.

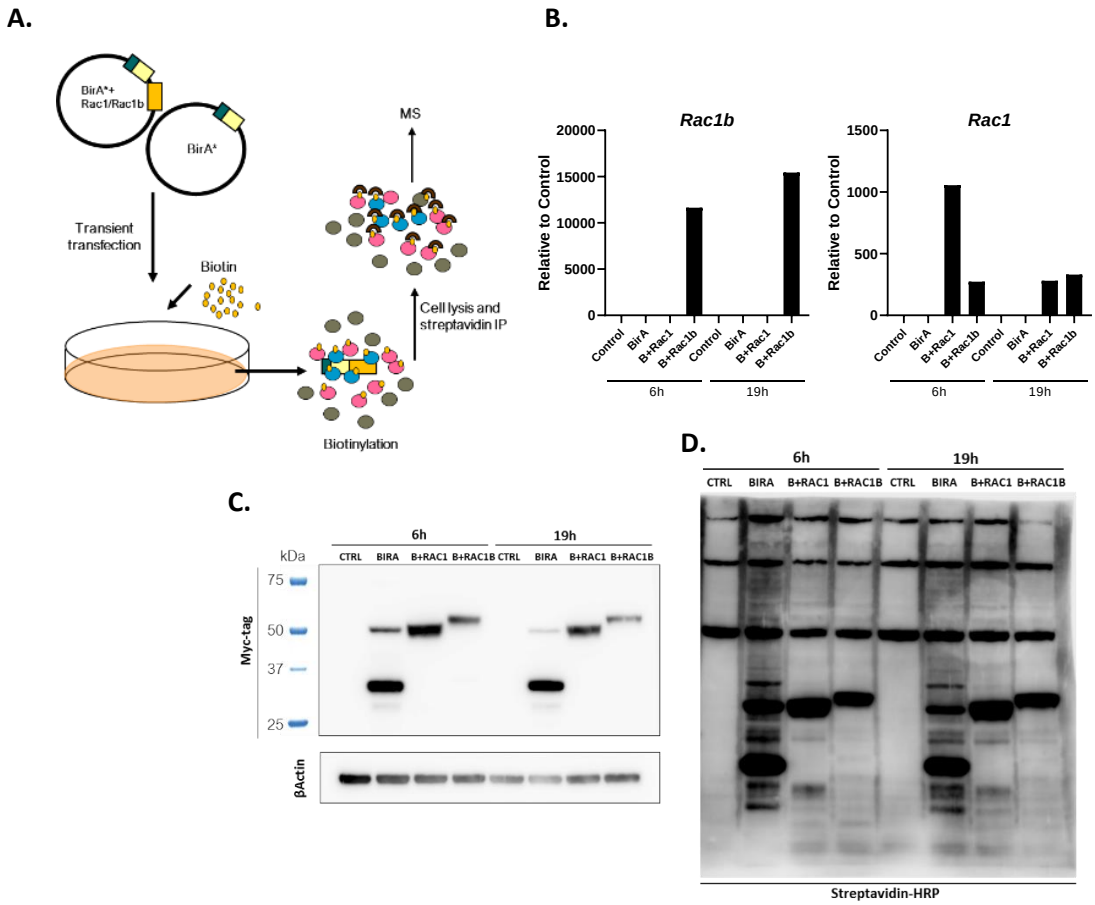


Figure 4.5: Set up of BioID experiment in the CMT93 cell line. A: Layout of the BioiD experiment using BirA*+Rac1 or BirA*+Rac1b plasmids and the non-fused BirA* plasmid as an experimental control. B and C: Validation of Rac1b and Rac1 overexpression by qRT-PCR (B) and WB probing for the myc-tag (C). D: Streptavidin WB to identify biotinylated proteins, conducted on cells after 6h and 19h of biotin treatment. N (condition) = 1.

4.2.4.4 Mass Spectrometry analysis

Upon vector expression and biotinylation validation in the cell lines, I proceeded to the second phase of the BioID experiment. 20 hours of biotin treatment was established as the ideal length of time for detecting enough biotinylated proteins without background excess. Cells were collected and proteins were extracted in RIPA buffer supplemented with phosphatase and protease inhibitors. A small amount of lysate was used to validate biotinylation and the remaining was used for streptavidin pull-down. Due to the high affinity between streptavidin and biotin, biotinylated proteins were efficiently separated from non-biotinylated by streptavidin beads immersed in an agarose resin. Briefly, protein lysate was incubated in a rotator with a small amount of agarose for 6 hours at 4ºC. The reason for using an agarose-containing beads instead of magnetic beads was mainly to avoid interference with MS analysis[465], as well as being easier to handle. Following incubation, washing of the beads from unbound biotinylation-free proteins was achieved through centrifugations and sequential pellet washes. The final pellet was then sent to the MS facility (IGMM) for protein identification.

A list of 60 proteins specific for either Bir-Rac1 or Bir-Rac1b came up from the MS analysis (Figure 4.6A, Table 4.1 and Appendix 4). The first main conclusion from the experiment was that both Rac1 and Rac1b share the vast majority of their interacting partners, since only two proteins are specific for Rac1 (Stxbp6 and Sms) and 8 for Rac1b (Itga3, Rap1gds1, Fam171a1, Wasf2, Flrt3, Pak4, Ttll3 and Ccdc127). In order to visualise the PPI networks Rac1 and Rac1b proteins were involved in, a meta-analysis of the listed proteins was performed using the NetworkAnalyst web-based platform[466,467] and the IMEx interactome database (Figure 4.6C). Seven protein subnetworks resulted from the analysis, with Rac1 as the core protein in the larger and main subnetwork and validating the interactome as Rac1/Rac1b-mediated. Bearing in mind the membrane cellular localisation of Rac1 (and for Rac1b as suggested and discussed in chapter 3), it functions as a further validation of the fact that most of the interacting proteins have a membrane localisation (either receptors or transporters) (Table 4.1) and both membrane and cytoskeleton processes are among the top most significantly represented pathways (Figure 4.6B). Disappointingly, there were no direct core proteins from the WNT signalling pathway in the list, although some of the candidates could be indirectly related such as Cyclin-Y[468] or Prominin1[469]. Intriguingly, the major subnetworks were represented by the core proteins

Ephrin-B2, Erbb2, EGFR and IGF1R, all four implicated in major cancer-related processes such as ephrin[470], ERBB family receptors[471,472] and insulin growth factor[473] signalling pathways. These pathways are associated with cell growth processes, they are of huge interest in CRC and their yet undescribed connection with Rac1b confers an additional significance to these findings.

The ERBB family of receptors were the subnetwork of interacting proteins that most attracted our attention. Two of these receptors, Erbb2 and EGFR, were identified as proteins that both Rac1 and Rac1b interact with, as well as other EGFR-related signalling components such as Yes1[474], Asap1[475,476] or Ptprj[477]. Moreover, crosstalk between EGFR and WNT pathway have been previously suggested[478,479]. For this reason, I decided to pursue the potential crosstalk between EGFR and Rac1b signalling. Initially, I tested whether EGFR was a specific Rac1b-biotinylated protein. The BioID experiment was repeated in BirA- and BirA-Rac1b-transfected cells by biotin treatment and streptavidin-dependent pull-down, and a Western blot was conducted against streptavidin-HRP and EGFR antibody. As it is shown in figure 4.6D, only the product from the pull-down is positive for streptavidin, validating the biotinylation. However, EGFR is exclusively seen in the Bir-Rac1b lane, demonstrating that EGFR and Rac1b are either interactors or neighbouring proteins.

Overall, BioID has demonstrated to be a useful and informative tool to investigate the interactome of a certain protein. The experiment with Rac1 and Rac1b as bait proteins showed a concise list of interacting proteins which will lead to a new perspective into their proteomic surrounding landscape.

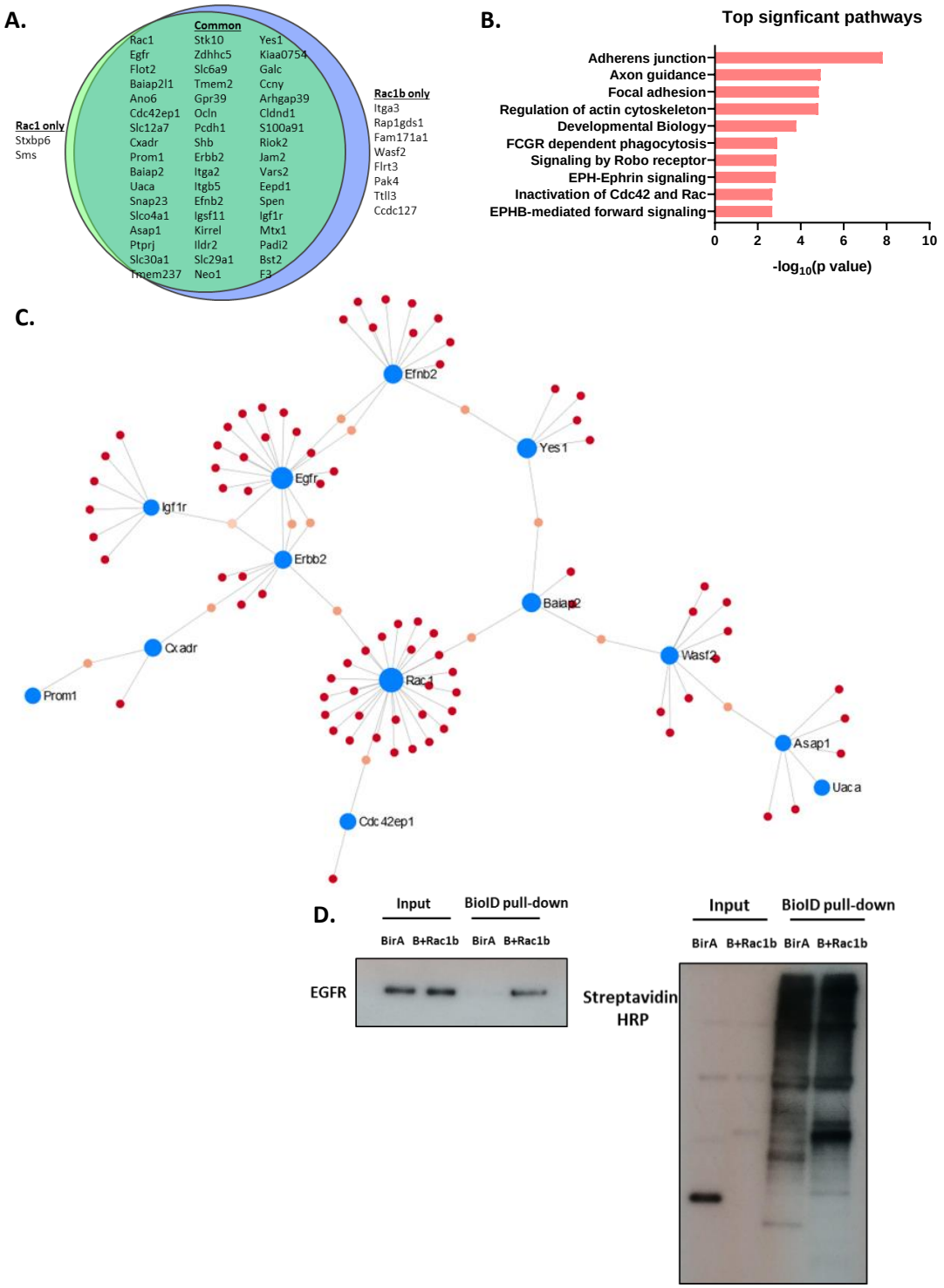


Figure 4.6: Members of the EGFR signalling pathway interact with Rac1b. A: 60 different proteins were identified as interactors of Rac1 and Rac1b, 50 of which were overlapped and 10 were specific for each one. B: Enriched signalling pathways in the BioID gene list, analysed by InnateDB. C: Main subnetwork identified for Rac1b interacting proteins, analysed by NetworkAnalyst. This include core members of oncogenic signalling pathways such as EGFR, Ephrin and Yes. D: Validation of EGFR as a Rac1b interacting protein by biotin-streptavidin-based pull-down. FCGR: Fcgamma receptor. MS analysis conducted by Alexandre von Kriegsheim. N= 3 vs 3 vs 3.

Gene name	Protein name	Interaction
Rac1;Rac2	Ras-related C3 botulinum toxin substrate 1;Ras-related C3 botulinum toxin substrate 2	Rac1/Rac1b
Itga3	Integrin alpha-3;Integrin alpha-3 heavy chain;Integrin alpha-3 light chain	Rac1b
Egfr	Epidermal growth factor receptor	Rac1/Rac1b
Flot2	Flotillin-2	Rac1/Rac1b
Baiap2l1	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1	Rac1/Rac1b
Ano6	Anoctamin-6	Rac1/Rac1b
Cdc42ep1	Cdc42 effector protein 1	Rac1/Rac1b
Slc12a7	Solute carrier family 12 member 7	Rac1/Rac1b
Cxadr	Coxsackievirus and adenovirus receptor homolog	Rac1/Rac1b
Prom1	Prominin-1	Rac1/Rac1b
Baiap2	Brain-specific angiogenesis inhibitor 1-associated protein 2	Rac1/Rac1b
Uaca	Uveal autoantigen with coiled-coil domains and ankyrin repeats	Rac1/Rac1b
Snap23	Synaptosomal-associated protein 23	Rac1/Rac1b
Rap1gds1		Rac1b
Slco4a1	Solute carrier organic anion transporter family member 4A1	Rac1/Rac1b
Asap1	Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 1	Rac1/Rac1b
Ptprj	Receptor-type tyrosine-protein phosphatase eta	Rac1/Rac1b
Slc30a1	Zinc transporter 1	Rac1/Rac1b
Fam171a1		Rac1b
Wasf2	Wiskott-Aldrich syndrome protein family member 2	Rac1b
Tmem237	Transmembrane protein 237	Rac1/Rac1b
Stk10	Serine/threonine-protein kinase 10	Rac1/Rac1b
Zdhhc5	Palmitoyltransferase ZDHHC5	Rac1/Rac1b
Flrt3		Rac1b
Slc6a9	Sodium- and chloride-dependent glycine transporter 1	Rac1/Rac1b
Tmem2	Transmembrane protein 2	Rac1/Rac1b
Gpr39	G-protein coupled receptor 39	Rac1/Rac1b
Ocln	Occludin	Rac1/Rac1b
Pcdh1		Rac1/Rac1b
Shb	SH2 domain-containing adapter protein B	Rac1/Rac1b
Erbb2	Receptor tyrosine-protein kinase erbB-2	Rac1/Rac1b
Itga2	Integrin alpha-2	Rac1/Rac1b
Itgb5	Integrin beta-5	Rac1/Rac1b
Efnb2	Ephrin-B2	Rac1/Rac1b
Igsf11	Immunoglobulin superfamily member 11	Rac1/Rac1b
Kirrel	Kin of IRRE-like protein 1	Rac1/Rac1b
Ildr2	Immunoglobulin-like domain-containing receptor 2	Rac1/Rac1b
Slc29a1	Equilibrative nucleoside transporter 1	Rac1/Rac1b
Neo1	Neogenin	Rac1/Rac1b
Pak4	Serine/threonine-protein kinase PAK 4	Rac1b
Kiaa0754	Uncharacterized protein KIAA0754	Rac1/Rac1b
Stxbp6	Syntaxin-binding protein 6	Rac1
Yes1	Tyrosine-protein kinase Yes	Rac1/Rac1b
Galc	Galactocerebrosidase	Rac1/Rac1b
Ccny	Cyclin-Y	Rac1/Rac1b
Arhgap39	Rho GTPase-activating protein 39	Rac1/Rac1b
Cldnd1	Claudin domain-containing protein 1	Rac1/Rac1b
F3	Tissue factor	Rac1/Rac1b
Riok2	Serine/threonine-protein kinase RIO2	Rac1/Rac1b
Jam2	Junctional adhesion molecule B	Rac1/Rac1b
Ttll3	Tubulin monoglycylase TTLL3	Rac1b
Vars2	Valine--tRNA ligase, mitochondrial	Rac1/Rac1b
Sms	Spermine synthase	Rac1
Eepd1	Endonuclease/exonuclease/phosphatase family domain-containing protein 1	Rac1/Rac1b
Spen	Msx2-interacting protein	Rac1/Rac1b
Igf1r	Insulin-like growth factor 1 receptor;Insulin-like growth factor 1 receptor alpha chain;Insulin-like growth factor 1 receptor beta chain	Rac1/Rac1b
		Rac1/Rac1b
Mtx1	Metaxin-1/Mitochondrial Outer Membrane Import Complex Protein 1	Rac1/Rac1b
Padi2	Protein-arginine deiminase type-2	Rac1/Rac1b
Bst2	Bone marrow stromal antigen 2	Rac1/Rac1b
		Rac1/Rac1b
S100a9	Protein S100-A9	Rac1/Rac1b
Cc+B3:B65dc127	Coiled-coil domain-containing protein 127	Rac1b

Table 4.1: List of BioID interacting proteins for Rac1/Rac1b. Highlighted rows correspond to proteins either specific for Rac1 (blue) or Rac1b (yellow).

4.2.5 Functional study of Rac1b-EGFR signalling pathway crosstalk

4.2.5.1 ERBB family receptors

Four different types of receptor tyrosine kinase (RTK) compose the ERBB or the Epidermal Growth Factors Receptor (EGFR) family: Erbb1 or EGFR, ErbB2 (or HER2), ErbB3 and ErbB4. In a ligand-free state, receptors are found inactive as monomers and they require homo- or heterodimerization for phosphorylation of tyrosine residues in the cytoplasmatic receptor tail[480]. Ligand binding to the receptor causes a conformational change that allows dimerization prior to receptor activation[481]. EGFR and ErbB4 are fully functional as homo- or heterodimers. In contrast, ErbB3 lacks intrinsic tyrosine kinase activity, albeit it still can transduce downstream signals through heterodimerization with the other ErbB receptors[482]. All of them can bind to a wide range of EGF-family of ligands except for ErbB2, which has only a fixed monomer constitution ready to form heterodimers with the other receptors. Although it has no ligand specificity, ErbB2 is the preferred receptor for heterodimer formation[483] and high concentrations of ErbB2 have been found in the cell membrane of many tumours, leading to increased EGFR signal activation.

Phosphorylated residues in the carboxy-terminal domain activate docking sites for recruiting proteins that will initiate intracellular signalling pathways[484]. Aberrant activation of ErbB receptors have been found in a large variety of human tumours. EGFR is the most mutated ErbB receptor family in epithelial cancers, usually by gene amplification or extracellular domain deletion[485]. It is followed by ErbB2, most commonly mutated in breast cancer[486]. Consequently, targeting the EGFR signalling pathway either by monoclonal antibodies or small-molecule inhibitors is currently a major therapeutic strategy[487]. Unfortunately, their secondary effects, low percentage of efficiency or acquired resistances demand more finely developed drugs or complementary treatments to improve the outcome of EGFR-driven tumours.

4.2.5.2 Kinetics of CMT93 cells following EGF stimulation

Regulation of EGFR signalling is mainly achieved by receptor recycling through the endosomes or lysosome targeting by ubiquitylation signals[482]. Once EGFR binds to EGF, its main ligand, the number of receptors in the cell surface diminishes and therefore, EGF is in turn its major negative feedback too. Even so, EGF-EGFR dimers are still active and continue

signalling after endosome internalization[488]. The EGFR signalling dynamic can be monitored depending on the EGFR peak of phosphorylation and downstream pathways activation. However, its kinetics will vary depending on the cellular context. Since I decided to use the CMT93 cell line to interrogate whether Rac1b is functionally required for EGFR downstream signalling, an initial characterization of EGF-stimulated EGFR activation in this cell line was needed.

EGF concentration is another factor to consider during *in vitro* cell stimulation[488]. Therefore, two different EGF concentrations were supplemented in the media following 4 hours of serum starvation. Upon EGF addition, cells were collected after 5, 10, 15 and 30 minutes and proteins were extracted for WB analysis. As it shown in figure 4.7A, EGFR phosphorylation peaked at 5 minutes and then the signal gradually decreased. MAPK and AKT pathways were checked as downstream activation read outs and both had the same pattern of activation like EGFR, with p-AKT and p-ERK1/2 reaching their maximum of activation at 5 minutes. Both EGF concentrations led to the same signal dynamic but 20ng/ml was much stronger and then more likely to detect differences if reduced, so this was the preferred concentration for the experiment.

4.2.5.3 Exon-3b targeted morpholino efficiently knock-outs Rac1b in CMT93 cells

The use of a targeted morpholino to Rac1 exon 3b was chosen as a strategy to knockdown *Rac1b* in the CMT93 cells. Antisense morpholino induce exon skipping by binding to exon splice enhancers and disrupting binding of splicing factors. There is evidence of promising therapeutic results with exon skipping morpholinos in diseases such as Duchenne muscular dystrophy and Spinal muscular atrophy[489,490]. For Rac1b, a specific morpholino was designed to bind to the exon 3b splice enhancer site thus blocking exon 3b inclusion. Morpholino knockdown efficacy was then tested in CMT93 cells. Cells were treated with 5µl/ml of either non-targeted (control or NT PMO) or targeted morpholino (*Rac1b* PMO) for 2 days and then lysed for RNA extraction. qRT-PCR analysis of three replicate experiments shows an approximate 95% of reduction in *Rac1b* transcript compared to NT PM ($p=0.0095$, Figure 4.7B). Importantly, *Rac1* transcript remained intact, indicating the morpholino's efficiency at only targeting the splicing event and not total transcript. Therefore, pre-

treatment of morpholino prior to the experiment is an efficient method to transiently knock down *Rac1b* expression in 2D cells.

4.2.5.4 Rac1b deletion impairs EGFR phosphorylation and downstream signalling activation

CMT93 cells were pre-treated with either NT PMO or Rac1b PMO for two days and then treated with 20ng/ml of EGF after serum starvation. Protein collection times were at 5, 10, 15 and 30 minutes and untreated EGF cells with 10% of serum or following serum starvation were used as controls. Interestingly, *Rac1b*-deleted cells presented a marked reduction in the phosphorylated status of EGFR, implying lower levels of activation (Figure 4.7C). This difference in p-EGFR was most evident at 10 minutes, and notably, *Rac1b* PMO treated cells never reached a level of phosphorylation as high as the NT PMO. Technical replicates of the experiment showed the same results and densitometry quantification of p-EGFR bands relative to their EGFR expression demonstrated a significant reduction in EGFR activation at 10 minutes (p=0.035).

Whether decreased EGFR phosphorylation translated into lower downstream EGFR pathway activation was assessed by ERK1/2 and AKT phosphorylation. As it is illustrated in figure 4.7D and E, both proteins were less phosphorylated in *Rac1b* PMO treated cells compared to controls. In this case, pathway activation levels were steady at all time points and so the reduction in phosphorylation observed in *Rac1b*-deficient cells was also equally diminished at the measured times.

Given that 10 minutes treatment with EGF displayed the most different time point of EGFR phosphorylation between cell conditions, concurrent replicates were performed in serum starved cells with EGF treatment for 10 minutes. Accordingly, *Rac1b* PMO treated cells presented a decreased activation in p-EGFR as well as in p-ERK1/2 and p-AKT, as evidenced by a 2-fold reduction in all phospho-proteins by densitometry analyses (Figure 4.7F).

These results strongly suggest that Rac1b modulates EGFR receptor and signalling activation. Intriguingly, the clearest difference in phosphorylation levels is after 10 minutes of EGF stimulation. At this time point, EGF-EGFR dimers are being internalised at the early endosome, suggesting that Rac1b might participate during EGFR recycling and therefore, it could alter the exposure and signal transmission time of the receptor-ligand dimer.

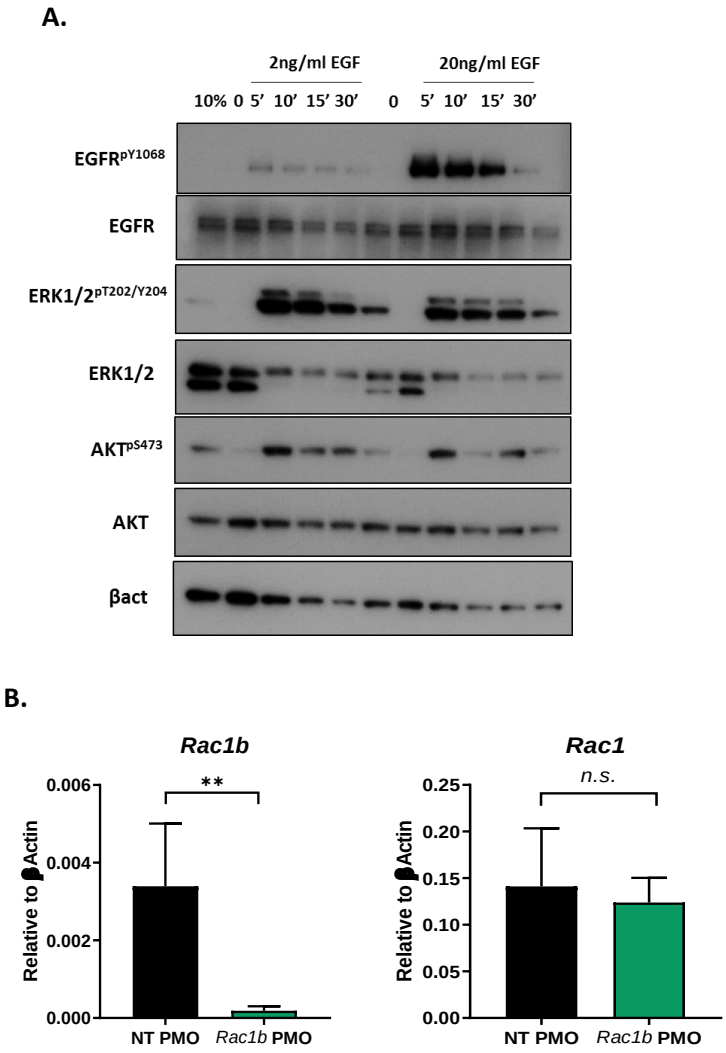
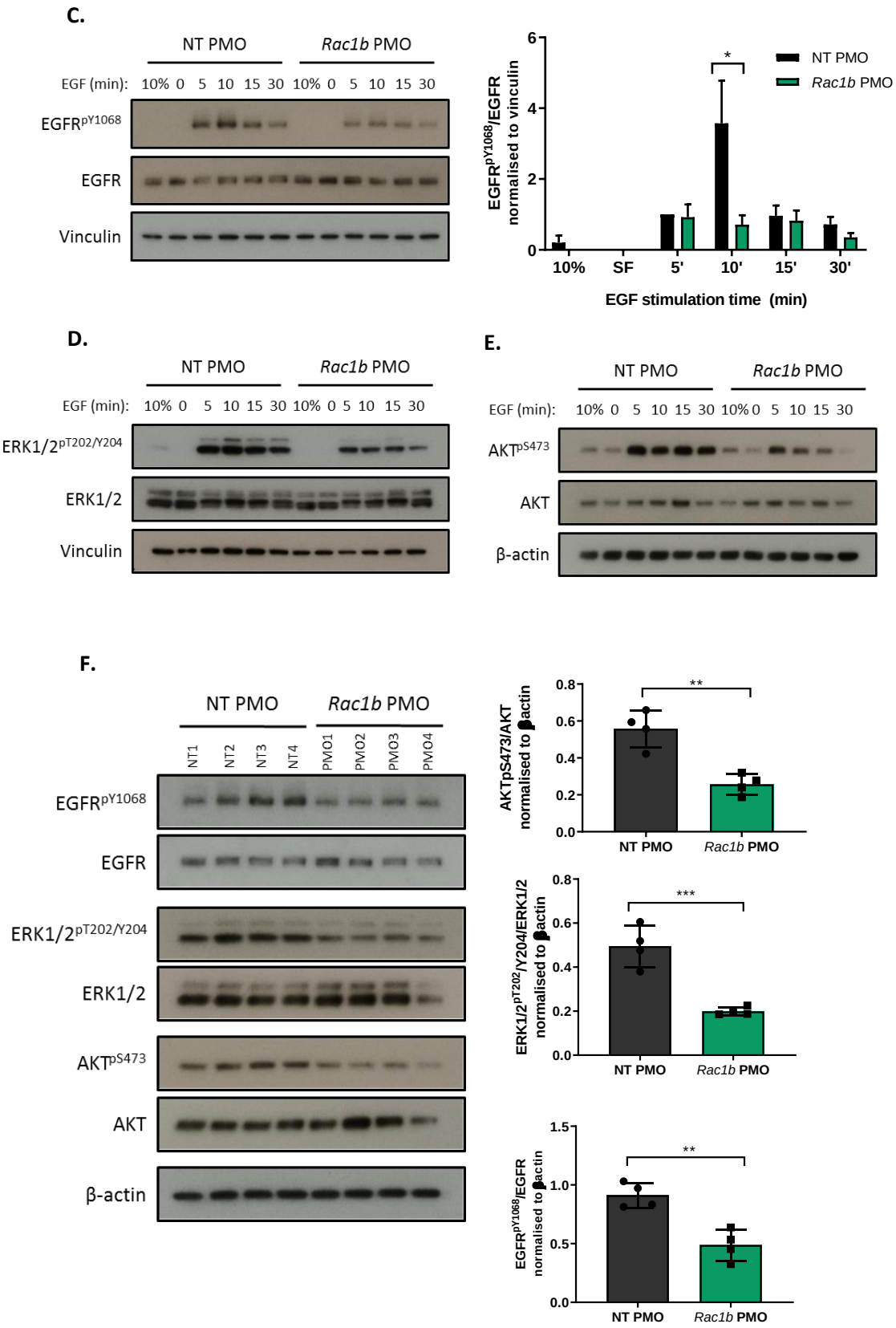


Figure 4.7: Deletion of *Rac1b* impairs EGFR phosphorylation and downstream signalling activation. A: Response of CMT93 cell line to EGF stimulation at two different doses, evaluated by EGFR, AKT and ERK phosphorylation. 20ng/ml was the chosen concentration to test EGFR signalling modulation in this cells. B: Validation of efficient *Rac1b* depletion by morpholino treatment on the CMT93 cells. qRT-PCR presents its downregulation, while *Rac1* expression remains intact. C-E: *Rac1b*-depleted (*Rac1b* PMO) and control (NT) cells following morpholino pre-treatment were subjected to EGF stimulation for 5, 10, 15 and 30min. Signalling activation was evaluated by EGFR (C), ERK1/2 (D) and AKT (E) phosphorylation. This experiment was conducted in triplicate. Band densitometries for p-EGFR are plotted in the adjacent graph, whose values are relative to non-phosphorylated EGFR. F: PMO and NT pre-treated cells were stimulated with EGF during 10 min. Four simultaneous technical replicates were conducted and EGFR downstream signalling was evaluated. EGFR, ERK and AKT phosphorylated proteins are quantified relative to their no phosphorylated expression.
T-test: *= p≤0.05, **= p≤0.01, ***= p≤0.001, n.s. = not significant, p>0.05. N (test) = 1 vs 1. N (time course) = 3 vs 3. N (replicate 10min) = 4 vs 4.


C.
NT PMO
Rac1b PMO
EGF (min): 10% 0 5 10 15 30 10% 0 5 10 15 30
EGFRpY1068
EGFR
Vinculin
EGFRpY1068/EGFR normalised to vinculin
NT PMO
Rac1b PMO
10% SF 5' 10' 15' 30'
EGF stimulation time (min)
D.
NT PMO
Rac1b PMO
EGF (min): 10% 0 5 10 15 30 10% 0 5 10 15 30
ERK1/2pT202/Y204
ERK1/2
Vinculin
E.
NT PMO
Rac1b PMO
EGF (min): 10% 0 5 10 15 30 10% 0 5 10 15 30
AKTpS473
AKT
β-actin
F.
NT PMO
Rac1b PMO
NT1 NT2 NT3 NT4 PMO1 PMO2 PMO3 PMO4
EGFRpY1068
EGFR
ERK1/2pT202/Y204
ERK1/2
AKTpS473
AKT
β-actin
AKTpS473/AKT normalised to βactin
ERK1/2pT202/Y204/ERK1/2 normalised to βactin
EGFRpY1068/EGFR normalised to βactin
NT PMO
Rac1b PMO

4.2.5.5 Organoids from VilCre^{ERT2}Apc$^{fl/fl}$Rac1b$^{fl/fl}$ mice present an EGF-sensitivity phenotype

4.2.5.5.1 Impaired cell growth with EGF withdrawal

Growing intestinal cells as organoids requires recapitulation in culture of the conditions that maintain the stem cell niche *in vivo*. Murine epithelial cells from the small intestine are probably the easiest to grow, since only three growth factors are needed to retain high activity of WNT and EGFR signalling and to inhibit BMPs: R-spondin-1, EGF and Noggin[491]. In the scenario of tumours or intestinal epithelial cells derived from acute hyperplasic tissues whose WNT signal is already hyperactivated, R-spondin-1 is unnecessary. However, EGF is essential for stem cells' normal maintenance and proliferation[151], whereas Noggin is dispensable for their expansion but needed for long-term culture[491].

We previously observed that organoids derived from *Apc* and *Rac1b* deficient mice have decreased clonogenic capacity compared to mice with wild type *Rac1b* expression (Figure 4.4F). Bearing in mind the EGFR modulation observed in CMT93 cells by Rac1b, we wondered whether these organoids would suffer differently upon EGF withdrawal. APC and APC RAC1B^{KO} organoids were split and an equal number of fragments were cultured for each group. Culture media was supplemented with either Noggin and EGF (+EGF) or just with Noggin (-EGF) and organoid size was evaluated. After four days, there was no difference in the size of organoids when EGF is present in growth media. However, when EGF was withdrawn, *Rac1b*-deficient organoids were significantly affected. More apoptotic organoids were observed in APC RAC1B^{KO} organoids without EGF and their size was significantly smaller than APC organoids (101,989 vs 57,503µm^3, p=0.0155, Figure 4.8A). Moreover, lysozyme WB from these organoids indicates that there was not an increase in the Paneth cell population to compensate for the lack of EGF (Figure 4.8C). This result indicates that organoids with *Rac1b* deletion have lower basal levels of EGF signalling than APC organoids and therefore, their survival might be compromised if EGF is no longer supplemented.

4.2.5.5.2 Increased sensitivity to EGFR inhibitor treatment

EGFR tyrosine kinase inhibitors (TKIs) treatment is a targeted therapy which is preferentially used in non-small-cell lung cancer (NSCLC), since a subset of these tumours harbour EGFR kinase domain mutations[492]. However, other cancer subtypes with high EGFR

pathway activation also benefit from this therapy such as cholangiocarcinoma[493], breast[494] and colorectal cancer[495,496]. Although efficient, a shared problem of this treatment is the development of resistance, which leads to surviving tumoral cells, tumour recurrence and death within 1- or 2-years[497]. The response of non-canonical pathways downstream EGFR activation to EGFRi have been explored as a possible tumour vulnerability, whereby their concomitant targeting would more efficiently lead to drug-sensitivity in tumour cells. A postulated one is the Rac1 signalling pathway, which is activated following EGFR phosphorylation and contributes to cell proliferation and migration signals via MAPK and c-Jun cascades[498,499]. In erlotinib-resistant gliobastoma cells, *Rac1* expression was shown to be associated with erlotinib sensitivity[500] and double treatment with erlotinib and Rac1 inhibitor (NSC23766) proved to act synergistically and efficiently enhanced their anti-neoplastic effects *in vitro* and *in vivo* in a xenograft mouse model[501]. A similar effect was observed in NSCLC-resistant cells, where NSC23766 and gefitinib combination reduced cell migration and growth[502]. Indeed, a recent work from Marcar and colleagues has demonstrated the effect of a Rac1-related drug against a large spectrum of TKI-resistant lung tumoral cells[503]. They showed that unlike TKI-sensitive cells, resistant cells produce more amounts of Rac1-derived ROS species, which makes them more dependent of a poly (ADP-ribose) polymerase 1 (PARP-1), indispensable to catalyse Rac1 and delimit their ROS cytotoxicity. This is a promising strategy of synthetic lethality against resistant tumours and PARP inhibitors are now under initial phases of a clinical trial. However, none of these studies have explored the possibility of Rac1b being involved in EGFR-triggered Rac1 activation. Given that *Rac1b* is its constitutively active tumoral isoform and bearing in mind the mounting evidence of EGFR-Rac1 signalling combination, I decided to investigate whether the tumour-like organoids derived from APC and APC RAC1B^{KO} short-term induced mice would respond differently to EGFRi treatment.

APC and APC RAC1B^{KO} organoids were treated with 1µM of an EGFR small molecule inhibitor (PD153035, Sigma®)[504] or its vehicle control (DMSO). After 20 hours, pictures were taken for organoid size measurement and resazurin was added in the media for 7 hours as a cell viability assay. As can be seen in Figure 4.8B, APC RAC1B^{KO} organoids treated with PD153015 performed noticeably worse compared to APC controls, since most of the organoids were dead or shrunk. While both genotypes remained the same when treated with DMSO, area analysis revealed that *Rac1b*-deficient organoids were 2-fold smaller in size upon EGFRi treatment, with almost no effect in control APC organoids (p=0.0037). Resazurin

absorbance measurements showed that cell growth was affected in both genotypes but APC RAC1B^{KO} organoids had a significantly worse viability compared to APC-treated organoids (p=0.048).

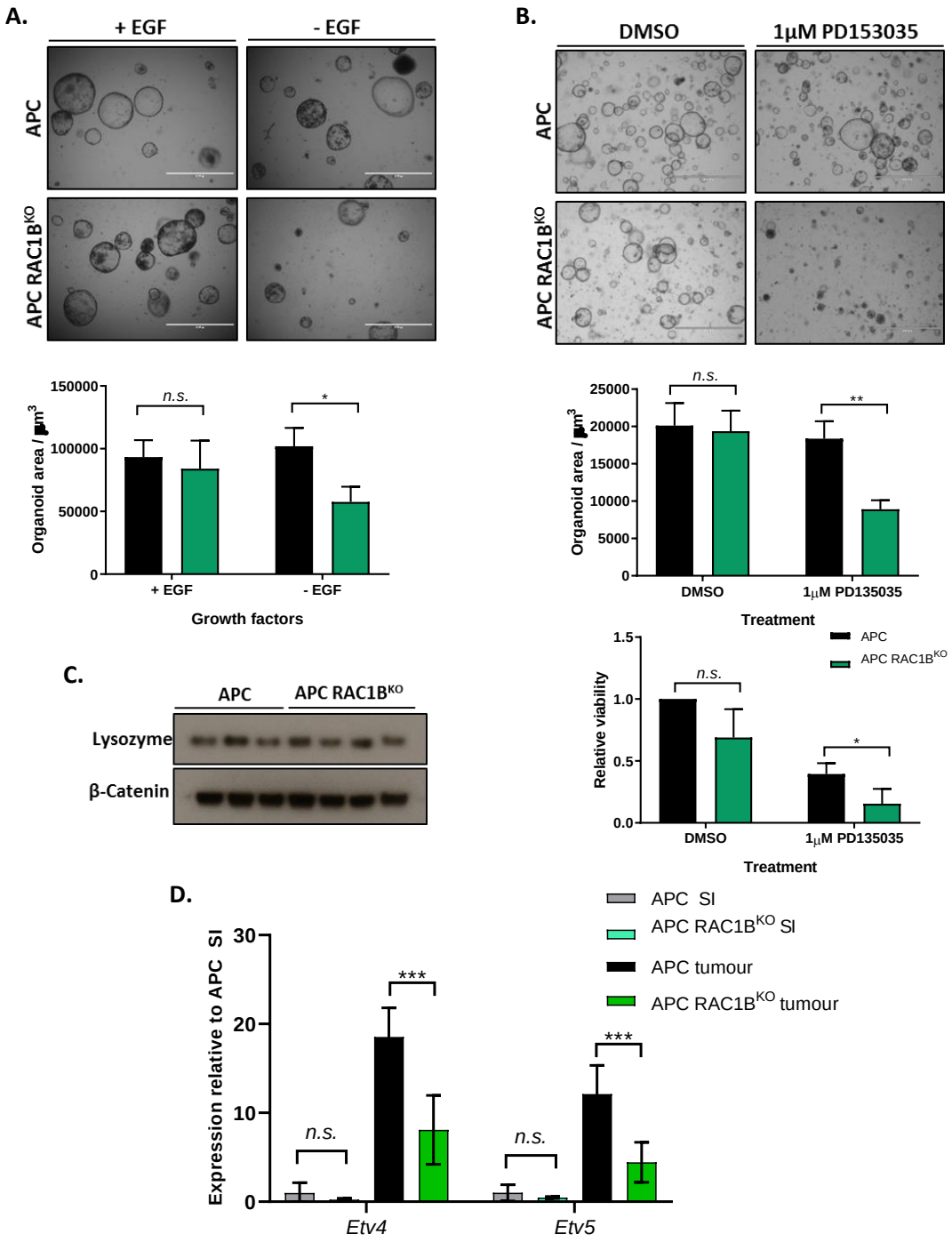
Altogether, these results demonstrate that deletion of *Rac1b* in adenoma-like organoids with high WNT activation are more sensitive to EGFR inhibition and suggest reduction of both signalling pathways as an attractive therapeutic strategy in high WNT activated CRC tumours.

4.4.3.1 EGFR signalling pathway activation is decreased in Rac1b-deficient mice

The ageing cohort presented a marked WNT-driven phenotype, evaluated by RNAseq analysis and qRT-PCR validation which might be responsible for both the downregulated cell proliferation within the tumours and the consequent decrease in tumorigenesis and extended survival in *Rac1b*-deficient mice. However, interactome analysis pointed to an additional yet unexplored signalling network led by the ERBB tyrosine receptors family. *In vitro* 2D and 3D analysis revealed a modulation of EGFR signalling by *Rac1b* expression, translating into a worse cell growth and an increased sensitivity to TKI treatment. Whether the same regulation occurs *in vivo* and contributes to tumour formation remains to be elucidated.

To this end, tumour and matching normal tissue samples from the ageing cohort were subjected to transcriptome analysis of the EGFR pathway through qRT-PCR analysis. *Etv4* and *Etv5*, two transcriptional target genes of the ERK pathway, were highly significantly downregulated in *Rac1b*-knock-out tumours compared to APC (p=0.0005 and p=0.0007 respectively, Figure 4.8D). Besides, *Etv4* already came up as one of the most downregulated genes in the RNAseq, validating its depletion upon *Rac1b* deletion.

Altogether, this result indicates that, in addition to the WNT signalling, deletion of *Rac1b* causes a decrease in EGFR signalling and both pathways might be responsible for the reduced tumorigenesis seen in this model.



<u>Figure 4.8</u>: *Ex-vivo* and *in vivo* modulation of EGFR signalling by *Rac1b* expression. A: APC and APC RAC1B^{KO} tumour organoids pictures with or without EGF-supplemented growth media. B: Pictures of tumour organoids upon EGFRi (PD135035) treatment. Size of the organoids for both experiments (A and B) was scored manually with ImageJ. Cell viability for PD135035 treatment was assessed with Resazurin. C: WB presenting lysozyme expression of tumour organoids. D: qRT-PCR of tumour and matching normal tissue from cohort mice for *Etv4* and *Etv5* expression. Experiments were conducted alongside Kevin Myant.
T-test: *= p≤0.05, **= p≤0.01, ***= p≤0.001, n.s. = not significant, p>0.05. N (organoids) = 3 vs 3. N (WB and qRT-PCR) = 4 vs 4.

4.2.6 Discussion

Since it was discovered in 1999, Rac1b has largely been described as a tumour-related isoform[385,391,395,505,506]. As important as it is to fully investigate its tumorigenic role, a comprehensive study of its function in physiological conditions is also necessary. My analyses indicate that Rac1b is not essential for normal intestine homeostasis as its deletion did not translate into any physiological change (Figure 4.1). This is not entirely surprising, given its low levels of expression. However, it adds extra value to the suggested tumour specificity for Rac1b and provides evidence for a lack of potential secondary effect in normal tissue upon artificial deletion. Furthermore, in this section I have presented how *Rac1b* contributes to tumour formation in a mouse adenoma model and investigated which signalling pathways were modulated and constitute its mechanisms of action. By unbiased transcriptomic and proteomic analyses, two major signalling networks with broad implications in cancer-related pathways were found to be altered upon *Rac1b* deficiency, and their modulation was studied by means of functional *in vitro* techniques.

In vivo deletion of *Rac1b* in cancer animal models has not been reported yet and given the building evidence of cancer-associated functions for Rac1b, an urgent need to conduct *in vivo* loss of function experiments in tumour models was apparent. Interestingly, *Rac1b*-deficient mice had an extended survival due to a decreased tumorigenesis, which was mainly developed in the most proximal parts of the intestine (Figure 4.2).

Phenotypically, none of the cell populations previously observed to differ upon modulation of Rac1b expression were different between genotypes[409] but interestingly, APC RAC1B^{KO} tumours had a diminished proliferation rate (Figure 4.3). RNAseq analysis from these tumours and additional qRT-PCR from acutely induced mice revealed the WNT signalling pathway to be significantly downregulated in *Rac1b*-deficient tumours (Figure 4.4). This correlates well with the reduction in proliferation since WNT is the major signalling pathway orchestrating proliferation and cell division processes in the intestine[142]. Moreover, a clonogenicity assay of intestinal epithelial cells derived from acutely induced *Apc* knockout mice presented a reduction in their clonogenic capacity, an *in vitro* phenotype analogue to that of reduced proliferation in tumours and reduced stem function. It needs to be mentioned that nuclear β-catenin, which is the main histological read out for WNT activation, is not decreased in APC RAC1B^{KO}, even though our transcriptome analysis shows that WNT-targeted genes are downregulated. This also indicates that Rac1b mediated regulation of the

WNT pathway is likely downstream of β-catenin nuclear localization, in agreement with the phenotype observed following *Rac1* deletion[292].

A correlation of WNT signalling activity and high Rac1b levels was also observed in my TCGA analysis from the previous chapter, where human colonic tumours with high *Rac1b* expression positively correlated with both the CMS2 subtype (the canonical WNT-activated subtype) and with WNT-targeted genes such as *Lgr5*, *Axin2*, *Tcf7* and *Lef1* (Figure 3.7). The fact that our own RNAseq data and GSEA analysis overlapped with the results observed in a large set of human tumours strongly supports the hypothesis that Rac1b is a functional mediator of WNT signalling. This is not completely unexpected, since Esufali *et al.* previously suggested a model, albeit controversial, whereby Rac1 and Rac1b activate WNT signalling through nuclear targeting of β-catenin[507]. Later on, they presented an upstream Dishevelled-3-dependent (Dvl-3) mechanism through which Rac1b would enhance β-catenin/Tcf-mediated transcription in the nucleus which in turn is WNT3A ligand independent[387,505]. However, additional evidence and the present observation that Rac1b likely acts downstream β-catenin signalling suggests certain discrepancies about Rac1 and Rac1b being necessary for its nuclear localization[292].

Considering the Rac1b-driven WNT signalling modulation, I was encouraged to investigate its mechanism of action by a PPI assay using BioID. Ensuing Mass Spectrometry analysis resulted in a concise list of interacting or neighbouring proteins, 60 in total, 50 of which were shared between Rac1 and Rac1b (Figure 4.6). This large overlap in their interactomes suggests that Rac1 and Rac1b might likely share most of their effectors, thereby Rac1b would enhance Rac1 activity rather than activate specific downstream signalling. Most of them were membrane proteins such as receptors (IGFR, EGFR, Cxadr and Gpr39), transporters (Slc12a7, Tmem237 and Slc29a1) or cell adhesion proteins (Ocln, Flot2 and Jam2), acting as evidence for Rac1/Rac1b preferential membrane localization. Protein meta-analysis revealed three branches of subnetworks that come out from Rac1 protein (Figure 11B). One is linked with Rho GTPases-related proteins with Cdc42 effector protein 1 (Cdc42ep1) as the core of the subnetwork. Another one is associated with Rac1 effector proteins involved in membrane ruffling and lamellipodia formation (Baiap2 and Wasf2)[508,509] and endocytosis (Asap1 and Uaca)[510,511]. Finally, the last branch connects to the major set of subnetworks, the majority of which are cell receptors related with cell growth processes and involved in pathways deregulated during cancer, such as the ErbB family of receptors (EGFR,

ERBB) and Insulin growth factor 1 receptor (IGF1R). A few proteins of the list could have a direct implication with the WNT signalling pathway, such as Prominin1 (CD133), which is considered an intestinal stem cell marker and a WNT regulator[469,512], and Cyclin-Y, which phosphorylates Lrp6 and induces activation of WNT/β-catenin pathway in a cell cycle-dependent manner[468]. I initially tried to study CMT93-morpholino treated cells response to the WNT ligand WNT3a, since theoretically, they have no mutations in the WNT pathway[438]. However, stimulated and non-stimulated cells responded equally to WNT3a treatment, evaluated by β-catenin WB and WNT-targeted genes qRT-PCR (data not shown), implying mutations somewhere in the pathway and requiring a different cell model to functionally investigate WNT modulation by Rac1b. An indirect relationship with the WNT pathway could also occur through other signalling pathways, such as EGFR[513,514]. Two out of four receptors from the ErbB family came up in the MS analysis, including other proteins that have already been proven to be related with EGFR signalling (such as Yes1, Asap1 and Prprj). A recent study by Wang *et al.* provided evidence for an EGF-Rac1b axis, since EGF would induce both ubiquitylation of the negative Rac1b-splicing regulator hnRNPA1 and activation of the SRPK1/2 and, consequently, increasing *Rac1b* expression[398]. In agreement with my data, this indicates a possible feedback regulatory loop between EGFR and Rac1b. However, a potential EGFR-modulation by Rac1b has not been described nor has any mechanism of action. With this rationale, the CMT93 cell line was used with an exon-3b targeting or control antisense morpholino to explore if Rac1b interaction with EGFR-related proteins translates into a functional phenotype.

Interestingly, EGF treatment to CMT93 cells showed a different activation of the pathway depending on *Rac1b* expression. Monitoring EGFR activation demonstrated that, after 10 minutes of EGF supplementation in the media, there was the largest difference between cell conditions in all evaluated proteins. It is believed that at this time point, EGFR-EGF dimers will be in the early endosomes sorting to multivesicular bodies (MVB) or late endosomes[515], suggesting that a depletion of *Rac1b* could have caused alterations in the endocytic route.

Prior to constructing any hypothetical model, we wanted to validate our findings in additional models that more closely reassemble human conditions[491]. To this end, organoids derived from tamoxifen-induced mice with double homozygous deletion in both *Apc* and *Rac1b* and their *Apc* knockout controls were plated and growth was evaluated under EGF-

deficient conditions. EGF is an essential growth factor in organoid culture for maintenance and proliferation of stem cells[151] and withdrawing it from the media would subject the organoids to extra stress. However, Paneth cells in the intestinal crypts are capable of producing EGF[136] and they could partially and temporally overcome the lack of EGF. Organoid size measurement following EGF starvation treatment revealed that only *Rac1b*-deficient organoids had an impaired growth, since they were visibly more affected and phenotypically, their size was reduced (Figure 4.8A). Western blot of these organoids showed that they did not have fewer Paneth cells, as previous analysis from ageing tumour cohort (Figure 4.8C) and normal tissue depleted for *Rac1b* has already revealed (Figure 4.1). Therefore, less intrinsic production of EGF due to a decreased Paneth cell population does not account for the EGF increased sensitivity in APC RAC1B^{KO} organoids and, more likely, a diminished activation of the EGFR signalling pathway as observed in the CMT93 cells might be responsible for their dependency.

I did not have time to fully interrogate EGFR signalling output in the ageing tumour cohort, but initial analysis showed the expression of ERK-targeted genes *Etv4* and *Etv5* was downregulated in Rac1b deficient tumours (Figure 4.8D). This suggests that *Rac1b* expression is also important for EGFR signalling *in vivo* and it acts predominantly in tumour cells, making it a potential therapeutic target.

Upon EGFR ligand binding and activation, a broad spectrum of signalling pathways are activated downstream and contribute to their malignant phenotype in tumour cells. Among them is the Rac1 signalling pathway, whose activation has already been investigated as a synergetic mechanism whereby EGFRi-treated tumours could be more efficiently eradicated[500–503]. Resistance to EGFRi is a growing problem, since a rising high percentage of tumours with elevated EGFR signalling, which initially have a good response to TKI drugs, will eventually develop resistance. An emerging therapeutic strategy is the combination of drugs which act at different levels the tumour is dependent on and weaken it before resistances are acquired. Given our results from CMT93 and tumour organoids, we hypothesised that efficiency of EGFRi could be increased if *Rac1b* expression is simultaneously deleted. To address this issue, APC and APC RAC1B^{KO} organoids were treated with the EGFRi PD153035 and cell size and viability was evaluated. *Apc* wild type organoids remained unaltered by the inhibitor, which is the optimal outcome for any cancer-targeting treatment. Interestingly, only *Rac1b*-depleted organoids were sensitive to the inhibitor (Figure 4.8C), demonstrating

that coupled deletion of *Rac1b* to an EGFRi could potentially enhance its outcome without affecting non-tumour cells.

Altogether, these were promising and novel results for Rac1b-related roles, and due to limited time, further description on its mechanisms of action could not be conducted. Observation of these data plus the data obtained from our RNAseq and MS analysis prompted us to postulate a hypothesis of how Rac1b could be interfering during EGFR signalling. As mentioned previously, the second branch of the subnetwork was related with membrane cellular processes, which are the canonical roles for Rho small GTPases and Rac1 in regulating the organization of the cytoskeleton, such as membrane ruffles, filopodia and lamellipodia. Rac1 modulates these processes by stimulating actin polymerization at the plasma membrane, whose dynamics have also been recognized to be important for membrane trafficking and endocytosis[516].

Endocytosis is the uptake mechanism of molecules from the surface and subsequent intracellular sorting for either degradation through multivesicular bodies (MVB) and lysosomes or recycling back to the plasma membrane. Upon ligand binding to membrane receptors, endocytosis has a prevalent role as a regulator and pathway attenuator of the signalling triggered by activated receptor tyrosine kinases (RTK)[517] and EGFR internalization serves as a working model for other RTK endocytosis. Briefly, upon EGFR dimerization and activation of phosphorylated binding sites, adaptor protein Grb2 and ubiquitin ligase Cbl are recruited, along other proteins, and start mono or polyubiquitylation of the receptor dimer. Although not necessary for receptor internalization[518], ubiquitin cues are recognized by ubiquitin binding proteins of the clathrin-coated pits which excise vesicle from the plasma membrane to an early endosome. Clathrin-independent endocytosis (CIE) have also been reported for EGFR internalization[519], although it is a slower mechanism and apparently only relevant during high EGF concentrations conditions, when clahtrin-dependent endocytosis (CME) might be saturated. At the endosomes, ubiquitylation can be recognized by the endosomal sorting complex required for transport (ESCRT) machinery which will target receptor dimer to lysosomes for degradation. Alternatively, deubiquitylating enzymes (DUBs) from the endosome plus other mechanisms can direct receptors to the recycling route and avoid degradation. In cancer, impaired endocytosis and imbalanced receptor recycling is considered an oncogenic signal, since altered dynamics in the endocytic traffic are associated with prolonged activation of tumorigenic signalling pathways[520].

In normal situations, EGF-bound EGFR is found at the early endosomes 2 to 5 minutes upon ligand binding and it is not until 15 to 20 minutes later when they translocate to the MVB and are sorted to the lysosomal pathway. In the CMT93 experiment wherein EGFR phosphorylation was decreased in *Rac1b*-deficient cells, the maximum difference was seen at 10 minutes after stimulation. At this point, receptors could still be recycled back to the membrane and it would suggest that Rac1b might somehow interfere in this step. A literature search of some of the candidate Rac1b-interacting proteins from the BioID experiment indicates that they have already been described to play a role during EGFR internalization or signalling, and their aberrant function assigns them as cancer drivers (Table 4.2 for an overview). On one hand there is Flotilin-1, which initially was shown to be essential for receptor dimerization[521] but recently has also been described to be involved in ESCRT internal movement and EGFR degradation[522]. On the other hand there is Ptprj or DEP1, a protein tyrosine phosphatase considered a tumour suppressor gene since it deactivates phosphorylated RTKs and it was found to inhibit EGFR signalling by physical interaction[477]. In addition, there is also Wiscott-Aldrich syndrome protein family member 2 (Wasf2 or WAVE2), which belongs to a complex in charge of mediating vesicle membrane fission, and recently, shown to take part in endosomal recycle to the plasma membrane through Arp2/3 complex activation[523–525]. Indeed, Rac1 has already been shown to indirectly activate WAVE proteins and to participate during vesicle movement by inducing actin polymerization[516]. This indirect activation is conducted by a protein which is also present in my MS list, the Baiap2 or ISRp53 protein[526]. Finally, another interesting candidate is Asap1. Asap1 is an Arf GAP protein that induces hydrolysis of the ADP-ribosylation factor (Arf) proteins, which recruit coat proteins and regulate vesicular transport[527]. Importantly, it has a BAR domain, as IRSp53 does[508], through which it mediates membrane bending and accelerates EGFR recycling[528]. The literature associates ASAP1 expression with tumour invasion and metastasis in a spectrum of tumour types[529–531] and poor human CRC survival[532]. Aside from modifying endosome recycling through its BAR domain[533], other mechanisms of action have been described. Asap1 is an effector of Arf6, whose activity promotes invasion[530,534]. Furthermore, Asap1 is also an effector of the adaptor protein CIN85 which, through its SH3 domain, binds to the ubiquitin ligase Cbl and decreases EGFR ubiquitination, increasing in turn its recycle pathway[510,531]. Interestingly, both mechanisms are equally appealing for the endocytosis related Rac1b hypothesis. Regarding CIN85, a GEF protein has recently been described which is a Cdc42 effector, Cool-1, whose overactivation in tumours displaces Cbl in an equal manner as CIN85

and consequently increases EGFR signalling[258], reaffirming CIN85-ASAP1 function in potentiating EGFR activation. However, a direct relationship between Arf6 and Rac1b regulating endocytic recycle in neurons has already been shown, where membrane trafficking and lamellipodia plays an essential role for neurons growth[535]. Altogether, this literature overview argues for a mechanism of action whereby Rac1b could modify EGFR signalling through regulating endosomal trafficking and likely, shifting its internalization to the recycle route instead of entry into the lysosomal pathway.

Gene name	Protein name	EGFR-related function	Reference
Egfr	**Epidermal growth factor receptor**		
Flot2	Flotillin-2	Endosomal sorting receptor degradation	Amaddi et al, 2012, The Journal of Biological Chemistry.
Baiap2l1	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1	Promotion of proliferation through EGFR-ERK signalling activation	Wang et al, 2013, Cancer letters.
Ano6	Anoctamin-6	Decreases ERK/AKT signalling. Direct EGFR interaction with Ano1 and Ano9	Zhao et al, 2014, PloS ONE.
Prom1	Prominin-1	Upregulates and stabilises EGFR in the cell surface to avoid receptor internalization	Wang et al, 2017, Cancer letters.
Baiap2	Brain-specific angiogenesis inhibitor 1-associated protein 2	Connects activated Rac and the WAVE complex to form lamellipodia	Suetsugu et al, 2006, The Journal of Cell Biology.
Snap23	Synaptosomal-associated protein 23	Invadopodia formation for EGFR-mediated invasion	Williams and Coppolino, 2014, Journal of Cell Science
Asap1	Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 1	Oncogene, accelerates EGFR recycling and avoids receptor degradation	Nie et al, 2006, Current biology.
Ptprj	Receptor-type tyrosine-protein phosphatase eta	Dephosphorylates EGFR for its inactivation	Tarcic et al, 2009, Current Biology.
Wasf2	Wiskott-Aldrich syndrome protein family member 2	Rac1-mediatic endosomal recycle	Jia et al, 2010, PNAS.
Stk10	Serine/threonine-protein kinase 10	Erlotinib off-target gene	Yamamoto et al, 2011, Molecular Pharmacology.
Flrt3		Phosphorylates ERK and activates signal in Xenopus laevis	Böttcher et al, 2004, Nature Cell Biology.
Gpr39	G-protein coupled receptor 39	Interact with β-arrestatins to activate Src and transactivate EGFR	Santos-zas et al, 2016, Cellular and Molecular Life Sciences.
Ocln	Occludin	EGFR activation regulates occludin levels in endothelial brain cells	Chen et al, 2011, Hepatology.
Erbb2	Receptor tyrosine-protein kinase erbB-2	Enhances EGFR activation through heterodimerization	Li et al, 2012, The Journal of Biological Chemistry.
Itga2	Integrin alpha-2	Related with erlotinib-induced resistance	Wu X et al, 2006, Genes and Development.
Itgb5	Integrin beta-5	Related with EGFR-TKI-induced resistance	Jacobsen et al, 2017, Nature Communications.
Efnb2	Ephrin-B2	Crosstalk with EGFR signalling	Bhatia et al, Clinical Cancer Research, 2018.
Pak4	Serine/threonine-protein kinase PAK 4	EGFR-PAK signalling axis for macropinocytosis in tumours, relevant in nutrients uptake	Lee et al, 2019, Developmental Cell-
Yes1	Tyrosine-protein kinase Yes	Activating insertions causes resistance to EGFRi	Fan et al, 2018, PNAS.
Riok2	Serine/threonine-protein kinase RIO2	Mediates EGFR downstream activation to AKT signalling	Read et al, 2013, PLoS Genetics.
Igf1r	Insulin-like growth factor 1 receptor;	Crosstalk with EGFR signalling	Van der Veeken et al, 2009, Current cancer drug targets.

Table 4.2: Rac1b/Rac1 interacting partners related with EGFR signalling. BioID targets with a described function in either processes related to EGF receptor recycling or involved in EGFR signalling pathway. Short description and literature reference are provided.

Nevertheless, I initially observed a strong WNT-related phenotype in the ageing cohort and its RNAseq analysis, and the Rac1b-WNT signalling association is further validated with the TCGA data. WNT and EGFR (or its downstream effector, Ras) signalling are major pathways maintaining tissue homeostasis and their aberrant activation have been detected in almost all cancer types[450,536]. Although their signalling mechanisms have been a topic of interest for decades and we currently have a large understanding of their function, their interconnection and crosstalk is more poorly understood. The synergistic potential of activated K-Ras and WNT signalling pathways has been observed in mouse models, leading to invasive tumours[537,538], but the underlying mechanism is not yet defined. A link was suggested through Axin1-mediated phosphorylation of H-Ras, whereby GSK3β phosphorylation is essential for H-Ras binding to the ubiquitin β-TrCP and targeting its degradation[478]. This indicates that both β-catenin and H-Ras utilise the same mechanism for protein degradation, and aberrant WNT activation due to loss of the β-catenin destruction complex enhances H-Ras stabilisation too. In liver cells, Tan and colleagues demonstrated that EGFR is a direct target gene of β-catenin[539]. Likewise, EGF can also induce nuclear translocation of β-catenin and enhance TCF/LEF1 driven transcription through increased E-cadherin caveolar-mediated endocytosis and/or α- and β-catenin complex disruption[514,540,541]. Destabilization of β-catenin complex in the membrane through E-cadherin downregulation and cell-cell adhesion alteration is suited to my model, since Rac1b-dependent E-cadherin decrease has been previously suggested[505] and an involvement of Rac1b maintaining cell membrane structure is by now fairly plausible. However, during my PhD project, results regarding Rac1b-E-cadherin regulation have not been conclusive and from that point of view, I cannot either discard or support this hypothesis. Both pathways could also be linked through their shared target gene c-Myc[144,542], whose expression shown a downregulation trend in the murine intestine with *Rac1b* depletion (Figure 4.4). Moreover, a recent investigation in haematopoietic stem cells has shown that EGFR functions as a co-factor for WNT9a-driven signalling[479]. They demonstrate that upon ligand binding, EGFR phosphorylates the receptor Fzd9b and this is essential for signal activation, hence demonstrating a functional link between these two pathways. Of note, the mouse model I used had *Apc* deletion as a tumour initiating event. Although expression of Rac1b correlates with *Apc* mutations in human tumours (Figure 3.9) and more probably, Rac1b-driven WNT regulation would be observed likewise in other mutational landscape, it would be interesting to test this phenotype in an additional WNT-independent CRC mouse model.

Dual inhibition of both pathways has recently been suggested as a necessary mechanism to efficiently treat all tumour populations, including CSC. Contrary to the current assumption for WNT and EGFR/Ras pathway acting synergistically, a recent study from Zan and colleagues have demonstrated that the use of MEK inhibitors (MEKi) can activate the WNT pathway. This activation occurs through *Axin1* downregulation, which induced tumour 'stemness' and could increase chances of relapse[171]. MEKi are commonly used to overcome resistance to EGFRi inhibitors, acting downstream of the receptor and oncogenic RAS or RAF. However, the side effect of increasing WNT signalling leads to an increase in stem cell gene signature in the tumour, suggesting a cell dedifferentiation reprogramming that would fuel the tumour with highly proliferative cells and induce treatment relapse. Interestingly, EGFRi inhibitors do not cause this effect, likely because they are not as potent in inhibiting Ras signalling as MEKi. While Zan *et al.* suggested a co-treatment with MEKi and WNTi, these results indicate that dual inhibition of EGFR and Rac1b is potentially equally efficient in sensitising tumours cells. Moreover, EGFRi and Rac1b inhibitor would downregulate the WNT pathway instead of activating it, preventing an unexpected side effect.

One of the points remaining that needs to be addressed is the effect of Rac1b in cells harbouring K-Ras mutations. *K-RAS* mutational status is one of the few parameters taken into accounting for clinical decision-making, as patients with K-Ras mutation cannot benefit from EGFR monoclonal antibodies. Given the finding of Rac1b as a modulator of EGFR signalling, it would be of major relevance to test whether its inhibition could sensitise *K-RAS*^{mut} tumours to EGFRi treatment. In the TCGA analyses from the previous chapter, I observed that *RAC1B* expression significantly correlated with *K-RAS* mutation in tumours. This might indicate a certain cooperation and/or requirement for Rac1b in tumours presenting *K-RAS*^{mut}. CRCs with mutations in *K-RAS* accounts for 40% of all tumours. Therefore, investigating whether *Rac1b* inhibition sensitises *K-RAS*^{mut} cancers could translate into a large impact.

In summary, the work from this section has described how depletion of *Rac1b* could inhibit two important pathways in intestine homeostasis, the WNT and EGFR signalling pathway, and translate into a less tumorigenic phenotype. Moreover, I have postulated a hypothesis whereby Rac1b could be regulating EGFR signalling through its participation in endocytic recycling routes. Although WNT and EGFR crosstalk is still unclear, I have shown that deletion of *Rac1b* in combination with EGFR inhibition might be a potential therapeutic

strategy, and the concomitant WNT inhibition may play a role in enhancing its treatment efficiency.

4.2.7 Future work

The results obtained from this work and the hypothesis built around them left some unanswered questions. One of the next experiments planned was to test Rac1b-dependent sensitivity on human colorectal cancer organoids, a more reliable and advanced model for clinic translation. These experiments were performed by other members of the laboratory and led to some interesting results. As mouse and human Rac1 exon 3b differs in 6 nucleotides, a new morpholino for human *RAC1B* was designed. Two types of human organoids cell lines were used, one derived from a benign polyp and another from an invasive carcinoma. When Rac1b was depleted by addition of morpholino in the media, only the polyp showed a significant reduced clonogenic capacity while the invasive organoid line did not present differences. However, when morpholino was administered co-treated with the EGFRi PD153035, both cells lines showed a significant decrease in proliferation compared to controls. This validates our previous findings and highly suggest that dual inhibition of Rac1b and EGFRi, rather than their single treatment, could potentially be an efficient therapy for advanced adenocarcinomas.

Following this, other experiments are necessary to complete a comprehensive characterisation of EGFR-Rac1b interaction. Firstly, it would be interesting to test Rac1b-dependent EGFR modulation in organoids derived from patients with *K-RAS* mutation or with acquired resistance to EGFR inhibitors. At the same time, I would like to test the endocytic recycle-related hypothesis. Initially, it should be validated if Rac1b has a physical interaction with some of the most interesting targets such as Asap1, Flotilin-1 or Ptprj. If confirmed, loss and gain of function experiments should be conducted to determine if they phenocopy *Rac1b* depletion. Additionally, there are several techniques to monitor EGF-EGFR internalization dynamics as for example, adding FITC tagged into the media and track GFP movement and localization on cells. I briefly tried this before finishing the project but the technique was not optimised and the results were not reliable. A combination of all these experiments should provide confident data to confirm or discard my hypothesis that Rac1b functions in membrane trafficking.

Aside from EGFR, there were other interesting interacting candidates such IGFR or Cyclin-Y, whose functions could also be important for Rac1b function. Experiments analysing their interaction in CMT93 and other cell lines and the phenotypes driven by their loss or gain of function would be valuable when describing Rac1b function and gaining a complete understanding of the Rac1b interacting landscape.

Finally, it would be intriguing to conduct further experiments to try to gain a better understanding of the potential WNT and EGFR signalling crosstalk mediated by Rac1b.

Completion of these experiments would then provide important data for the Rac1b requirement during early tumour development in addition to a better understanding of its modulated pathways in cancer.

4.3 Acute ectopic overexpression of *Rac1b* in the murine intestine does not alter homeostasis

Rac1b overexpression is achieved by the insertion of a human cDNA copy of *Rac1b* (*hRac1b*) in the Rosa26 locus after a stop cassette flanked by *loxP* sites[408]. I crossed these mice to the VilCreER^{T2} model to generate cohorts of *VilCreER^{T2}Rosa26$^{+/+}$* (WT) and *VilCreER^{T2}Rosa26$^{lsl\text{-}Rac1b/lsl\text{-}Rac1b}$* (RAC1B^{OE}) mice. Upon Cre expression, the *lox-stop-lox* sequence is excised and *hRac1b* is expressed (Figure 2.1). WT and RAC1B^{OE} mice were induced with tamoxifen and analysed 5 days later to investigate the effects generated by high levels of *Rac1b* in the intestine. Initially, induction was validated by qRT-PCR from SI tissue, showing a strong overexpression of *hRac1b* accomplished after 5 days, without altering either *mRac1b* or *Rac1* expression (Figure 4.9A).

Acknowledging that *Rac1b* has a preferential crypt location could suggest an intestinal stem cell-related role for Rac1b, with potential functions as diverse as control of proliferation, stem cell maintenance and renewal and/or intestinal cell lineage regulation, among others. Using qRT-PCR I analysed the expression of a set of stem cell marker genes in the SI. *Lgr5* was slightly but significantly increased in the RAC1B^{OE} group, but neither *Slc14a1* nor *Olfm4* were changed (Figure 4.9C). Besides, measuring cell proliferation in the crypts by BrdU incorporation did not show any differences among groups (Figure 4.9B), indicating that cell growth is not altered by *Rac1b* overexpression, and that the small increase in *Lgr5* does not account for any gross phenotypic change. Furthermore, analyses of Paneth cells by lysozyme IHC and goblet cells by both Periodic Acid Schiff (PAS) did not present any significant differences, implying that these cell populations remained at their normal state following *Rac1b* overexpression.

Intestinal homeostasis is controlled in part by regulation of the WNT signalling pathway. Therefore, I determined whether *Rac1b* overexpression could impact its normal function by analysing expression of a selection of WNT target genes in the murine intestine. As it is shown in figure 4.9C, none of them (*Axin2*, *Lef1*, *Tcf7* and *Tiam1*) are changed in the RAC1B^{OE} group, pointing to normal levels of activity of the WNT signalling pathway. I repeated the fractionation experiment explained in the previous section with RAC1B^{OE} mice and saw the same pattern of expression in all tested genes (Figure 4.9D). Therefore, I conclude that short term overexpression of *Rac1b* in wild type mice does not cause any gross changes to intestinal function.

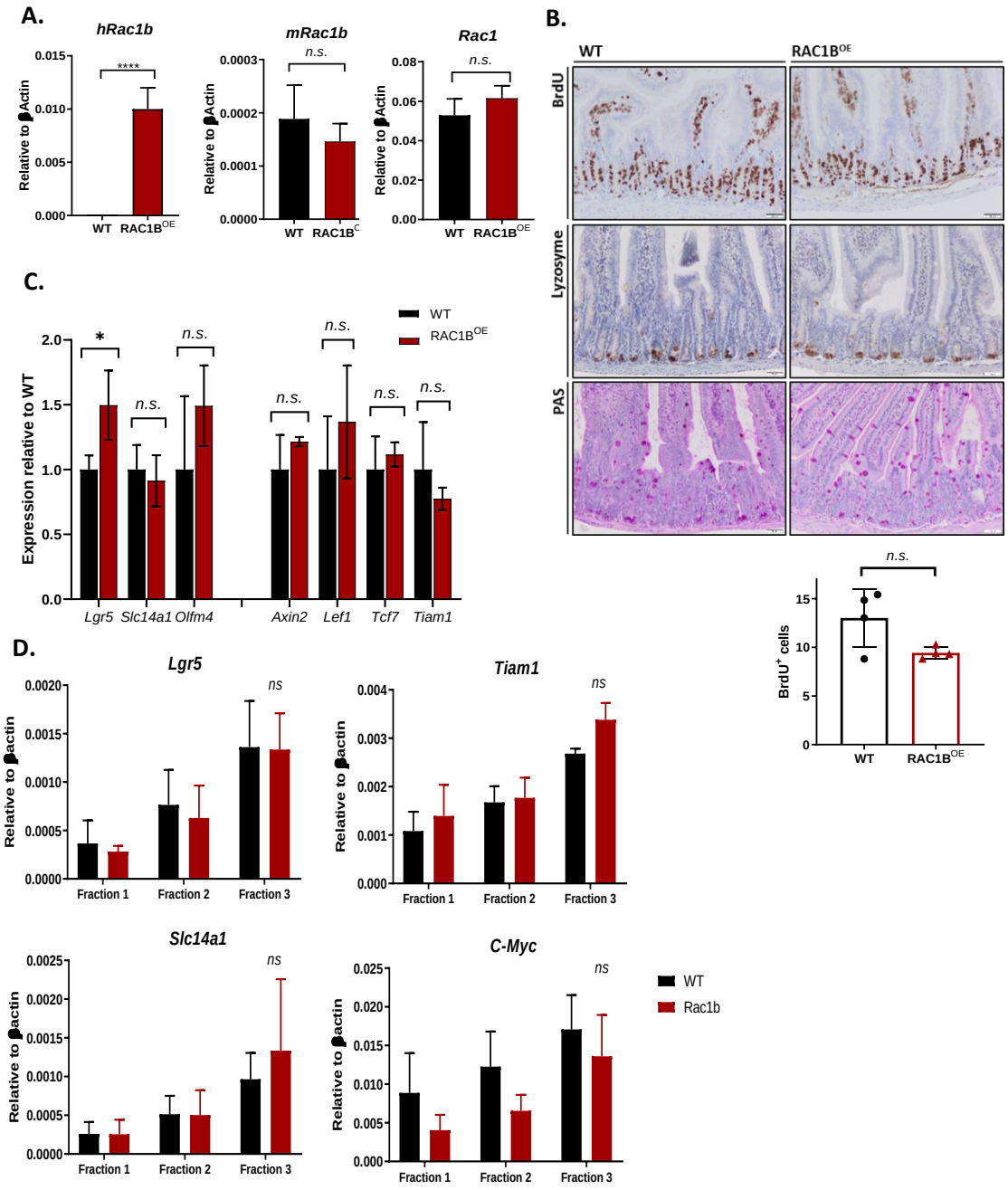


Figure 4.9: Overexpression of *Rac1b* in the normal intestine. A: qRT-PCR from SI tissue, showing that ectopic *Rac1b* is efficiently overexpressed without affecting endogenous *Rac1b* and *Rac1* expression. B: BrdU (proliferation), Lyzosyme (Paneth cells) and PAS (goblet cells) staining in the normal SI tissue. Proliferation score is plotted underneath, through manual score of at least 50 half crypts per mouse. C: qRT-PCR of normal SI tissue for stem cells markers (*Lgr5*, *Slc14a1*, *Olfm4*) and WNT/β-catenin target genes (*Axin2*, *Lef1*, *Tcf7*, *Tiam1*). D: qRT-PCR from fractionation experiment comparing the expression of stem cells (*Lgr5* and *Slc14a1*) and proliferation markers (*Tiam1*, *c-Myc*) in WT and RAC1B^{OE} tissue.
T-test: *= p≤0.05, ****= p≤0.0001, n.s. = not significant, p>0.05. N (qRT-PCR) = 3 vs 3. N (Histology) = 4 vs 4.

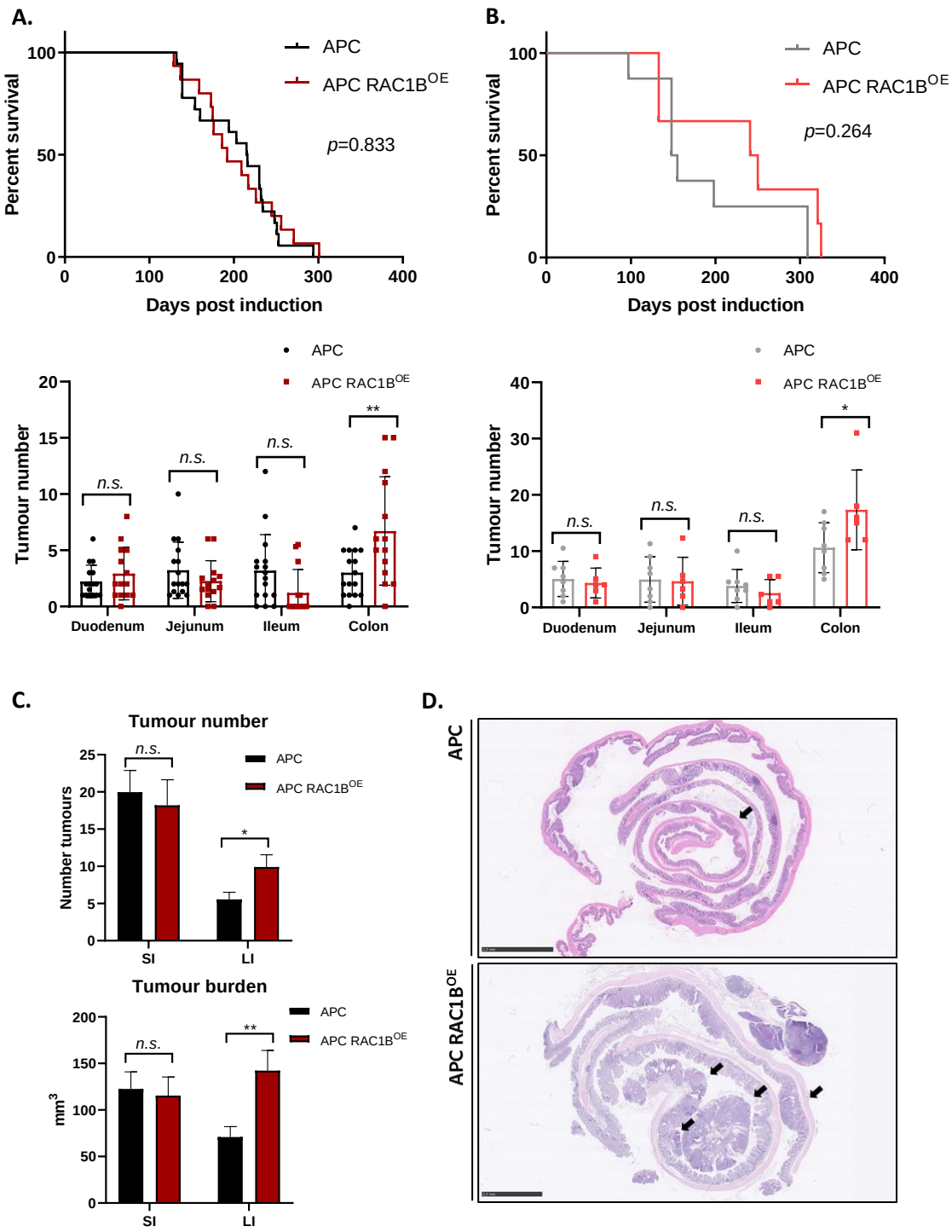
4.4 Overexpression of *Rac1b* increases colonic tumorigenesis

4.4.1 Ageing cohort with *Rac1b*-overexpressing mice

I have previously shown that *Rac1b* expression significantly increases in tumours compared to normal tissue, and this increase is higher in more advanced tumours from complex genetic models carrying multiple oncogenic mutations (Figure 3.11). Therefore, I sought to determine whether an ectopic overexpression of *Rac1b* in an adenoma could lead to a more aggressive carcinoma that might grow faster or even invade. Using the $VilCre^{ERT2}Apc^{+/-}$ and $Rosa26^{Rac1b/Rac1b}$ system, I studied two groups of mice: control $VilCre^{ERT2}Apc^{+/-}$ $Rosa26^{+/+}$ (hereafter APC or wild type) and $VilCre^{ERT2}Apc^{+/-}Rosa26^{Rac1b/Rac1b}$ (hereafter APC $RAC1B^{OE}$). Mice were tamoxifen induced and aged until they become apparent of disease (assessed by weight loss, feet paleness or presence of prolapses).

As it is shown in Figure 4.10A, intestinal overexpression of *Rac1b* did not cause a survival difference compared to the APC group (215.5 and 192 days for APC and APC $RAC1B^{OE}$ group). Accordingly, both genotypes developed the same number of tumours (18 and 22 tumours), indicating that higher levels of *Rac1b* did not enhance tumorigenesis and consequently, both groups presented the same survival. However, tumour analysis according to intestinal region revealed a significant difference in the number of colonic tumours. Mice with *Rac1b* overexpression developed twice as many tumours as the APC group (3 vs 6.7 tumours on average, p=0.0089, Figure 4.10A) which were also significantly larger (70.92mm^3 vs 141.2mm^3 respectively, p=0.0036, Figure 4.19C). Altogether, this suggest that *Rac1b* overexpression might promote tumorigenesis in the colon.

In order to validate this colonic phenotype, I repeated the same ageing cohort again but with smaller groups (n=8 and 6 for APC and APC $RAC1B^{OE}$, respectively). Interestingly, they performed the same as the previous cohort did: there were not differences either in survival or in overall tumour number, but APC $RAC1B^{OE}$ mice presented a significant increase in colonic tumour number (10 vs 15 tumours respectively, p=0.045, Figure 4.10B). Thus, overexpression of *Rac1b* in the intestine of mice upon *Apc* deletion facilitates tumorigenesis in a tissue-limited manner, leading to increased colonic tumour formation without altering overall survival.



<u>Figure 4.10</u>: *Rac1b* overexpression promotes colonic tumorigenesis. A: Data from the first ageing cohort induced mice, presenting Kaplain-Meier survival plot (Long-rank test) and tumour number based on intestinal tumour origin. Mice in the cohort (n): APC = 19, APC RAC1B^{OE} = 15. B: Same as (A) but data correspond to the second ageing cohort. Mice in the cohort (n): APC = 8, APC RAC1B^{OE} = 6. C: Tumour number and burden data merged from both cohorts. D: Swiss-roll overview of colons from cohort mice. Arrows mark tumours in the section.

T-test: *= p≤0.05, **= p≤0.01, n.s. = not significant, p>0.05.

4.4.2 *Rac1b*-driven colonic tumorigenesis does not appear to be mediated by WNT or EGFR signalling

In the previous section, I observed that deletion of *Rac1b* led to a decreased tumorigenesis and a reduction in proliferation of *Rac1b*-deficient tumours. Given that *Rac1b* overexpression consistently enhances tumorigenesis in the LI, both SI and colonic tumours were assessed for proliferative changes. Interestingly, while SI tumours present the same proliferative rate in both groups, APC RAC1B^{OE} colonic tumours displayed a slight but significant increase in proliferation (p=0.0426, Figure 4.11A). This indicates that Rac1b promotes colonic tumorigeneses through a specific increase in cell proliferation restricted to the LI.

As Rac1b appears to modulate WNT and EGFR signalling to promote tumorigenesis, assessment of these two pathways was conducted by qRT-PCR. Especially, I wanted to determine if a differential level of activation between SI and LI might explain the confined tumour promotion phenotype. RNA from SI and LI tumours and matching normal samples was extracted and cDNA was synthetized to conduct qRT-PCR analysis. Initially, I confirmed overexpression of *hRac1b*, which, as expected, is only expressed in samples from the APC RAC1B^{OE} group (Figure 4.11B and C). Subsequently, WNT target genes were evaluated. Even though the expression of all of them increased in SI and LI tumour samples, there were no differences between groups (Figure 4.11B and C). *Etv4* and *Etv5* were used as transcriptional markers for ERK1/2 activation, but none of the groups showed significant differences. Overall, this indicates that neither WNT or EGFR signalling are responsible for Rac1b-driven colonic tumorigenesis and different pathways might explain its molecular mechanism.

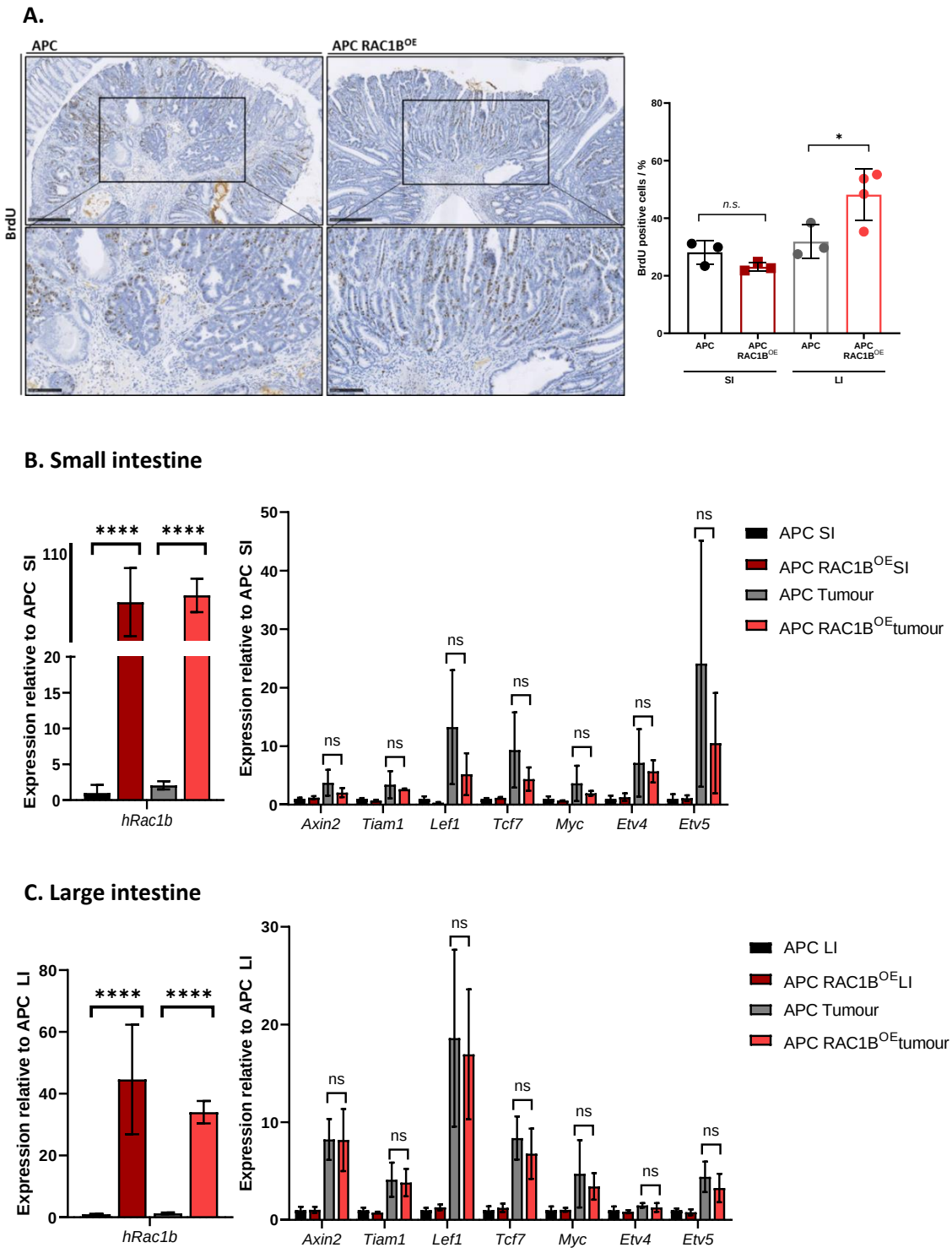


Figure 4.11: *Rac1b* promotes colonic tumour proliferation independently of WNT and EGFR signalling activation. A: Representative pictures of colonic tumours stained for BrdU IHC. Proliferation score was carried out by QuPath and percentage of proliferation is plotted in the adjacent graph. B and C: Validation of *Rac1b* overexpression and evaluation of the expression of WNT/β-catenin (*Axin2*, *Tiam1*, *Lef1*, *Tcf7* and *c-Myc*) and ERK1/2 transcriptional targets (*Etv4* and *Etv5*) by qRT-PCR from SI (B) and LI (C) tumours and matched normal tissue. T-test: *= p≤0.05, ****= p≤0.0001, n.s. = not significant, p>0.05. N = 3 vs 3.

4.4.3 Functional investigation of Rac1b in a time-point model

A caveat of studying samples from long experiments might be that, at their endpoint (about 200 days), *Rac1b*-driven processes that could point to a hint of its function might be masked by other co-occurring and additional genetic alterations acquired over time. Since I did not find major differences in the forecasted altered pathways, I sought an alternative approach to investigate this colonic-related phenotype. Even though acute overexpression of *Rac1b* did not translate into major homeostatic differences, it may be possible that, after a longer period of time, but prior to tumour initiation, activated pathways downstream of Rac1b could be observed. To this end, mice either wild type for *Rac1b* (WT) or overexpressing two copies of *hRac1b* (RAC1B^{OE}) were induced and terminated after 31 days (namely day-31 experiment). Unlike short-term experiments carried over 5 days, which can harbour collateral effects due to tamoxifen-dependent stem cell population depletion[543] or can miss slow-developing changes, I reasoned that any potential difference observed would exclusively be causative of *Rac1b* overexpression.

An initial characterization of the large intestines was conducted by both histology and qRT-PCR. As it is illustrated in figure 4.12A, there was no difference in their goblet cells -measured by PAS staining- nor in their inflammation status, evaluated by CD3 (T-cell lymphocytes marker) and F4/80 (macrophages marker) IHC. Moreover, qRT-PCR analysis of stem cells, WNT and ERK1/2 target genes showed no differences between groups (Figure 4.12C). Regarding proliferation, half crypts from the distal colon were scored and presented the same proliferative rate (Figure 4.12B). Therefore, this indicates that general intestinal homeostasis was not affected when *Rac1b* was overexpressed for 31 days in the murine intestine.

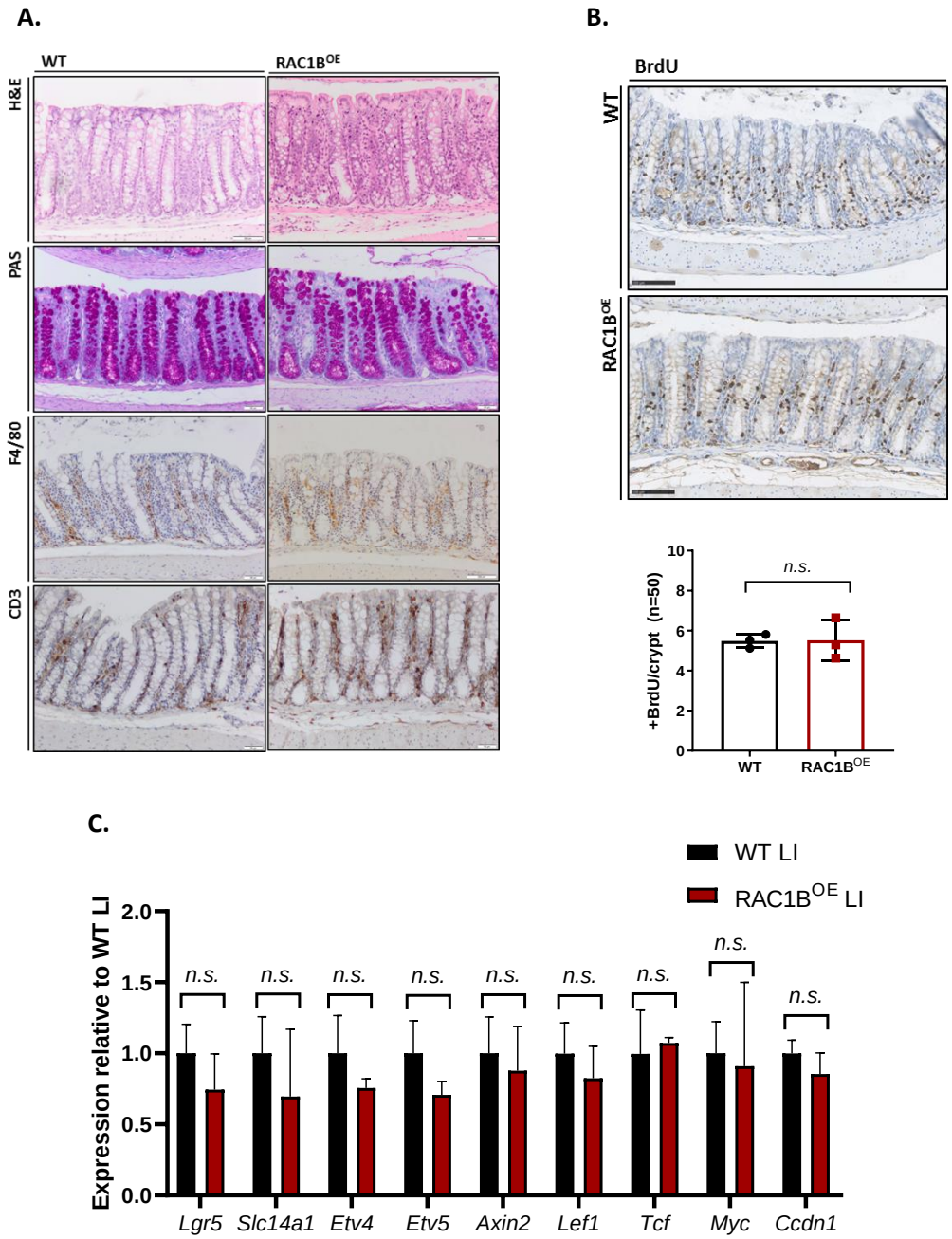


Figure 4.12: Short-term overexpression of *Rac1b* does not induce major changes in the murine colon. A: H&E, PAS (Goblet cells) F4/80 (macropaghes) and CD3 (T-cell lymphocytes) IHC from WT and RAC1B^{OE} colonic tissue. B: BrdU staining and proliferation score of from at least 50 half LI crypts. C: Evaluation of stem cell markers (*Lgr5*, *Slc14a1*), ERK1/2 (*Etv4* and *Etv5*) and WNT/β-catenin target genes (*Axin2*, *Lef1*, *Tcf*, *c-Myc* and *Ccdn1*) in colonic tissue from WT and RAC1B^{OE} mice by qRT-PCR.
T-test: n.s. = not significant, p>0.05. N = 3 vs 3.

4.4.3.1 Molecular evaluation of the day-31 model

Each intestinal region harbours many anatomical and functionally different traits. For instance, since nutrition absorption is done in the small intestine while the large intestine is in charge of water absorption and fermentation, villi structures are only found in the small intestine as a method to increase its surface area. Likewise, there is a large heterogeneity of commensal microbes (microbiome) depending on the intestinal region they reside, partially maintained by the asymmetrical Paneth cells distribution (restricted to the SI) or by distinct anti-microbial peptides the Enterocytes release along the intestine[544]. Different anatomy also translates into gradients of pathway activation and sensitivity to some diseases, such as inflammatory bowel disease (IBD) or cancer, which in humans is more frequent to develop in the large intestine than in the small intestine. Given the phenotype observed in the ageing cohort with *Rac1b* overexpression, an understanding of its differential function according to the intestinal region could be relevant in some colon-dependent human diseases. Taking advantage of the homogenous samples from the day 31-experiment, I proceed to investigate two clinically relevant cellular processes in relation to *Rac1b* overexpression: cell-adhesion and the TGFβ signalling pathway.

E-cadherin is a cell-cell adhesion molecule localised at the adherens junction region of epithelial cells[545]. Apart from being essential to organise proper cell contact, E-cadherin is also necessary for maintaining a ratio of E-cadherin:β-catenin complexes in the membrane. Therefore, loss of E-cadherin promotes cell detachment and invasion but, at the same time, it also leads to the release of β-catenin from the membrane and allow its translocation to the nucleus for initiating transcription of WNT target genes[546]. Work from Huels *et al.* showed that the murine LI expresses more E-cadherin compared to the SI, making the colon a more resistant region of the intestine in initiating tumour transformation. This helps explain why in most of the CRC mouse models, tumorigenesis more often occurs in the SI than in the LI, making it a more challenging translational model for studying human CRC. Given the suggested role for Rac1b in decreasing E-cadherin expression, I tested whether *E-cadherin* levels were altered upon *Rac1b* overexpression in the day-31 model. Initially, I performed qRT-PCR for *E-cadherin* in both SI and LI of WT and RAC1B^{OE} samples. As it is shown in figure 4.13A, the RAC1B^{OE} group has a 2-fold reduction in *E-cadherin* transcript compared to wild type mice specifically in the colon, whereas the SI has equally lower levels of *E-cadherin*

(p=0.0058). With the purpose of validating this result, I also looked at the protein level of E-cadherin by IHC. However, all six mice showed a quite similar pattern and intensity of staining (Figure 4.13B). IHC is not the most suitable technique to evaluate quantitative changes, but this indicates that, if any, Rac1b does not cause major protein changes in E-cadherin although the reduction in mRNA transcript warrants further validation.

Another signalling pathway with differential activation between small and large intestine is the TGF-β signalling pathway. This is a superfamily of proteins and receptors that are involved in many cellular processes such as cell proliferation and cell death, invasion and differentiation. Upon ligand binding and receptor activation, the canonical TGF-β signalling activates two different SMAD signalling cues that converge to associate with co-SMAD4 for nuclear translocation. SMAD4 is an essential cofactor for translating TGF-β signalling to the nucleus, and compelling evidence in the literature have shown the loss of *SMAD4* to be an important event during CRC development. Interestingly, a work from Means *et al.* has demonstrated that loss of SMAD4 in cooperation with an acute inflammation driven by Dextran sodium sulphate (DSS) was sufficient to initiate carcinogenesis in the colon of treated mice, emulating human colitis-associated carcinoma[547]. This finding, along with other similar publications[548,549], suggests that TGFβ might exert its tumour suppressor effect more efficiently in the large intestine compared to the small intestine. Given the suggested role for Rac1b in modulating TGF-β signalling, I decided to assess the activity of the pathway in my samples. Following a similar strategy as used to evaluate the potential Rac1b effect on E-cadherin levels, I looked at the status of the TGF-β pathway in the day-31 samples by qRT-PCR. Possible mechanisms Rac1b might use to inhibit the pathway could be downregulation of either SMAD4 or TGF-β receptors. Comparing RNA extracted of colonic tissue from WT and RAC1B^{OE} mice, I saw a small but highly significant decrease in SMAD4 transcript in RAC1B^{OE} compared to WT, and no difference in TGF-β receptor 2 (TβR2) (Figure 4.13C). *Bim*, a TGF-β target gene I have looked previously in other studies and has shown to have a promising link with Rac1b (further described in Chapter 5), is also downregulated at a similar fold change in the RAC1B^{OE} group, albeit not being significant. These results would initially point to a potential alteration of the pathway and suggest that Rac1b might reduce TGF-β activity in order to prevent its inhibition of proliferation.

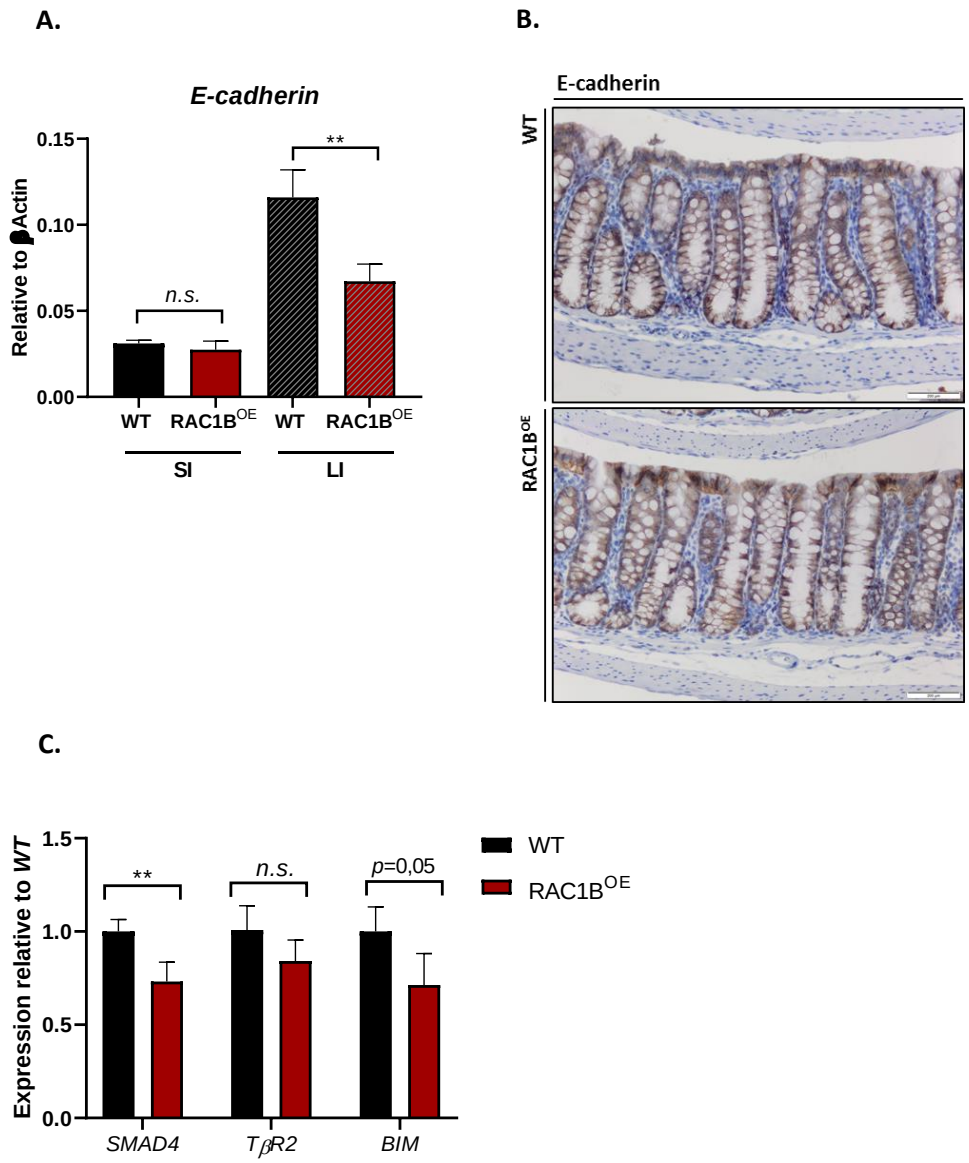


Figure 4.13: Potential downregulation of *E-cadherin* and *SMAD4* levels in the colon of *RAC1B* overexpressing mice. A: Assessment of *E-cadherin* expression in SI and LI tissue from WT and RAC1B^{OE} mice by qRT-PCR. B: E-cadherin IHC of colonic tissue. C: Colonic expression of *SMAD4*, *TβR2* and *Bim* measured by qRT-PCR.
T-test: *= p≤0.05, n.s. = not significant, p>0.05. N = 3 vs 3.

4.4.4 Discussion

In this section, I have provided information about the effects of ectopic overexpression in the murine intestine in both physiological and neoplastic conditions. On one side, I have demonstrated that overexpression of *Rac1b* does not translate into phenotypic differences in the intestine, neither after an acute induction for 5 days nor after 31 days of induction. This is contrary to the observations made by Kotelevets *et al.* who show, with the same model, that *Rac1b* overexpression causes increased proliferation, enhanced activation of the WNT signalling pathway and an expansion of Paneth cells[409]. However, their experimental mice were aged for 18 months, which is a substantially different experimental condition to my experiments and could lead to such discrepancies.

In order to assess whether Rac1b might potentiate the tumorigenic potential of the APC model, *VillinCRE^{ERT2}Apc$^{fl/-}$* mice were crossed with *Rac1b* overexpressing mice and were aged until symptomatic of disease for evaluating tumour number and survival differences. However, *Rac1b* overexpression did not reduce mice survival as it did not increase overall intestinal tumorigenesis. Intriguingly, it did increase colonic tumorigenesis reflected in both tumour number and size. This was consistently reproduced in two independent cohorts conducted at separate periods, validating these results. Interestingly, the increase in colonic tumours was not due to a shift from SI to LI tumorigenesis, since tumour number in the SI was unchanged. Survival changes in mouse CRC are predominantly driven by increased tumorigenesis in the SI, which explains the lack of differences in this model. Although mouse survival was not reduced, enhanced colonic tumorigenesis might imply that Rac1b cooperates with large intestinal traits not found in the small intestine and uncovering this mechanism could shed some light into processes important for colon-related diseases such as IBD or colitis-associated cancer (CAC).

Our first hypothesis was that this cohort resembled the opposite phenotype as seen in the *Rac1b*-knock-out cohort, but attenuated and restricted to the colon. Interestingly, cell growth analyses revealed that LI tumours, but not SI, had a slight but significant increase in proliferation. However, when WNT and EGFR signalling pathways were evaluated by qRT-PCR, neither of the samples presented any difference. Therefore, the current phenotype does not feature the opposite signalling outputs observed in the *Rac1b*-deficient mice, and new pathways should be investigated to understand its phenotype.

Given that I was analysing endpoint tumours, I was concerned that additional mutations acquired as a consequence of tumour formation could be covering and/or masking changes led by Rac1b. In order to circumvent these problems, a tumour-free cohort with or without *Rac1b* overexpression induced for 31 days was carried out. An initial histological characterization did not demonstrate major significant changes in any of the cell populations investigated. Next, I evaluated two molecular processes that have been already suggested to have both, a relation with Rac1b and a different output between SI and LI[505,550].

On one side, E-cadherin is a cell-adhesion protein that, on top of driving EMT and invasive processes, it serves as a buffer to keep β-catenin in the membrane. Given that the colon expresses higher levels of E-cadherin than the SI, I hypothesised that a reduction of E-cadherin in the LI reaches a threshold in the ratio of E-cadherin:β-catenin that enables tumour formation, which is smaller than the ratio needed for the same to occur in the SI. In the literature, there is controversy whether Rac1b is able or not to modify E-cadherin expression. Esufali *et al.* initially suggested a model whereby Rac1b reduces E-cadherin expression in HCT116 cells via stimulation of *Slug* (*Snai2*) expression, a negative regulator of E-cadherin and a WNT target[505]. However, just on year later, Lozano *et al.* argued that only Rac1 and Rac3, but not Rac1b, can alter E-cadherin localization and induce junction disassembly through their direct interaction with PAK1 in keratinocytes cells[551]. For the Rac1b colonic phenotype, it would be reasonable to think that Rac1b enables misslocalization of E-cadherin in the cell membrane, thereby disrupting cell-cell adhesion and favouring tumour initiation. Given the results from the pre-tumorigenic model and the evidence in the literature relating Rac1b and E-cadherin expression, this hypothesis warrants further investigations to providing a more definitive conclusion.

TGF-β is another signalling pathway that has shown activation differences across the intestine. Besides, a body of evidence from Ungefroren's lab and my own data (in the next chapter) have demonstrated a compelling link between Rac1b and TGF-β, specifically with SMAD4. SMAD4 is a tumour suppressor gene which is mutated in about 20% of all CRCs[183]. Interestingly, Freeman *et al*. demonstrated that loss of *SMAD4* in the context of *Apc* deletion led to enhanced WNT signalling pathway and a consequent increase in tumour number, albeit loss of *SMAD4* alone couldn't increase nuclear β-catenin[552]. Moreover, a recent work associated the loss of SMAD4 with colitis, providing evidence that low levels of TGF-β activity might facilitate tumorigenesis in an inflammatory context[547]. Interestingly, a link between

Rac1b and colitis has already been suggested by Jordan *et al*., who demonstrated that expression of *Rac1b* increases upon inflammation and its ibuprofen-mediated decreased expression reduces tumours originated from colitis-associated cancer (CAC). Bearing all this in mind, expression of *SMAD4* was tested in the day-31 and tumours samples and presented promising results which, as well as with E-cadherin, await further investigation.

Even though both pathways suggested promising results at the beginning, tumours could not validate these changes. Owing that the day-31 model was designed as a strategy to evaluate a cleaner cell population with just *Rac1b* overexpression, the observed differences have potentially been driven by Rac1b. However, tumours were the samples that showed a phenotypic difference and candidate downstream pathways should be equally reflected in those samples. Nevertheless, this study has been quite superficial and a deeper analysis should be carried out for any definite conclusion. Despite not showing differences in their survival, a permissive tumorigenesis in the colon might allow higher propensity to other diseases, especially those related with inflammatory processes. Hence further investigations might provide interesting insights with potential therapeutic relevance.

4.4.5 Future work

Due to time constraints of the project and overlapping with other interesting ongoing experiments, further investigations into this phenotype could not be carried out. However, deregulation of E-cadherin levels and/or downregulation of TGF-β signalling are feasible mechanisms to explain the observed phenotype and their analyses warrant further investigation. On one side, protein expression of E-cadherin should be analysed by Western blot in both day-31 colons and cohort tumours. Moreover, intestinal epithelial cells could be extracted from RAC1B^{OE} induced mice and functional experiments carried out. For instance, it would be interesting to fractionate their cellular content into membrane or cytoplasmatic fractions and measure E-cadherin expression by Western blot. Additionally, my previous BioID experiment demonstrated that Rac1b interacts with a number of cell adhesion proteins, such as Occludin or Jam2 (Junctional adhesion molecule B) (Table 4.1). An evaluation of their expression might indicate whether Rac1b caused cell membrane disruptions that facilitated tumorigenesis.

Regarding TGF-β signalling, pathway activity should also be evaluated at the protein level by WB. As well as measuring expression of SMAD4 based on *Rac1b* overexpression, functional experiments could also be conducted in *ex vivo* organoid culture of colonic epithelial cells from induced mice. TGF-β1 treatment of organoids induces activation of the pathway, which translates into induction of cell death and phosphorylation of its downstream effectors SMAD2 and 3. If Rac1b modulates TGF-β activity in the colon, its treatment with TGF-β1 ligand may reveal functional differences.

These experiments might help to clarify the potential implication of these two pathways. However, unbiased analysis of Rac1b differential activity in the SI and LI might explain more efficiently its mechanistic function, such as a RNAseq analysis comparing SI and LI with or without Rac1b overexpression or a BioID experiment comparing SI and LI Rac1b interactome. Although interesting, the study of Rac1b functions in other models yielded more striking phenotypic changes and therefore were the focus of my investigations.

Chapter 5: Function of Rac1b in tumour progression

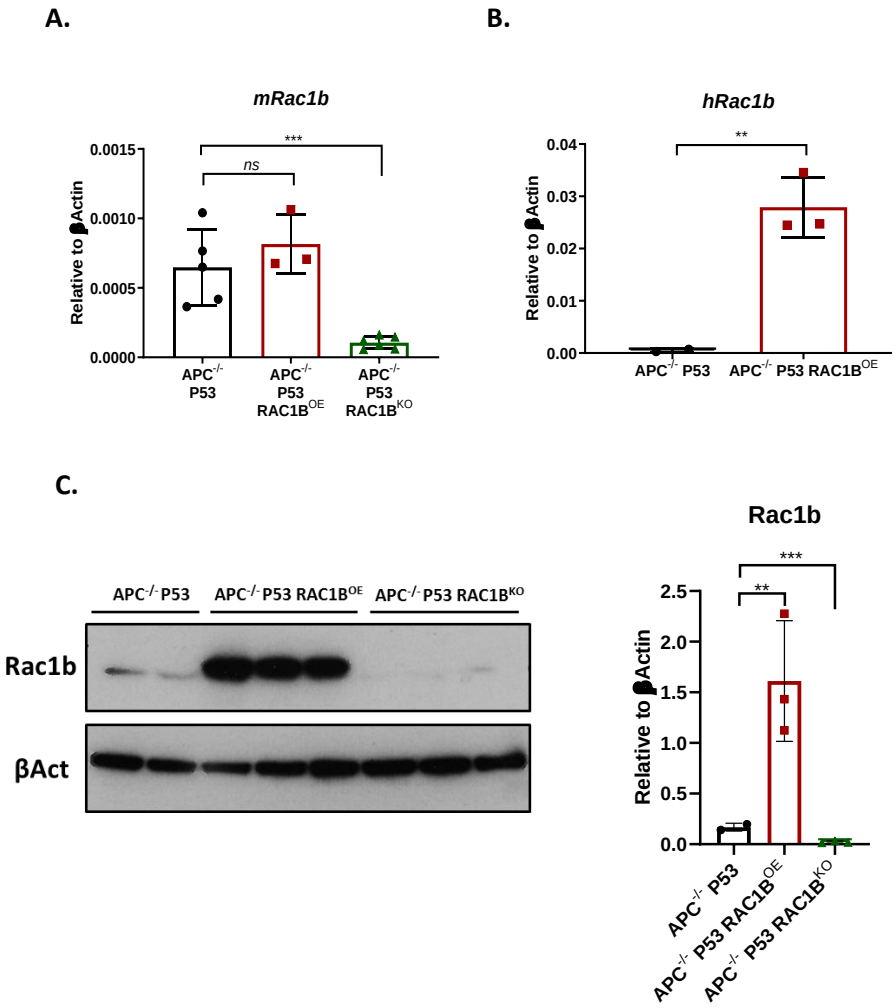
Mutations in the *TP53* gene occurs in all cancer types and its loss is implicated in leading CRC towards carcinoma progression[59]. Published literature and our own data have demonstrated that loss of *TP53* alone is insufficient to initiate intestinal tumorigenesis in mouse models[553]. However, *TP53* deficiency in cooperation with activated WNT signalling by either loss of *Apc* or by carcinogen (AOM)-induced carcinogenesis enhances tumorigenesis and leads to tumour invasion[195,553]. Preliminary analysis demonstrated that expression of *Rac1b* increases in *Apc* and *TP53* deficient mice compared to loss of *Apc* alone (Figure 3.11). Besides, my TCGA analysis showed a strong correlation between *RAC1B* expression and *TP53* mutations (Figure 3.9). Altogether, it suggests that conditional double deletion of *Apc* and *TP53* could be a suitable model to assess the potential invasive role of Rac1b. Therefore, *VillinCreER^{T2} Apc$^{fl/+}$ TP53$^{fl/fl}$* (APC P53) mice were crossed with either RAC1B^{KO} or RAC1B^{OE} mice. On one side, I aim to address whether Rac1b is required to drive intestinal tumour invasion. On the other side, by *Rac1b* overexpression I will investigate whether it can enhance the invasive potential of the model.

5.1 *Rac1b* deletion does not improve survival when co-deleted with *Apc* and *TP53*

Prior to initiation of the tumour ageing cohort, acute tamoxifen inductions were carried out in double homozygous *Apc* mice in order to validate the efficiency of *Rac1b* deletion in the model. *VillinCreER^{T2} Apc$^{fl/fl}$ TP53$^{fl/fl}$ Rac1b$^{+/+}$* (APC$^{-/-}$ P53) *or VillinCreER^{T2} Apc$^{fl/fl}$ TP53$^{fl/fl}$ Rac1b$^{fl/fl}$* (APC$^{-/-}$ P53 RAC1B^{KO}) mice were acutely induced for 4 days and tissues were harvested for molecular analysis. qRT-PCR and WB analysis from the small intestine demonstrated an efficient deletion of *Rac1b* both, at the RNA and protein level (Figure 5.1).

Therefore, similar to the previous cohort, two cohorts of mice were induced bearing the genotypes *VillinCreER^{T2} Apc$^{fl/+}$ TP53$^{fl/fl}$ Rac1b$^{+/+}$* (APC P53) or *VillinCreER^{T2} Apc$^{fl/+}$ TP53$^{fl/fl}$ Rac1b$^{fl/fl}$* (APC P53 RAC1B^{KO}). Following tamoxifen induction, symptom surveillance was carried by evaluation of pale feet and body weight. Mice started to become symptomatic of disease about 90 days after induction and termination was required shortly afterwards. As it is illustrated in figure 5.2A, both groups presented a virtually identical survival, 110 days for

APC P53 and 113 days for APC P53 RAC1B^{KO}. Accordingly, there was no difference on their tumorigenesis since APC P53 and APC P53 RAC1B^{KO} mice developed 74 and 59 tumours respectively (Figure 5.2B). Interestingly, none of the mice developed any invasive tumour and their termination was needed due to the number of tumours originated rather than their aggressivity. Thence, in this model, despite expectation, loss of *TP53* does not induce tumour invasion. This indicates that this model acts as a tumour initiation model rather than a tumour invasion and therefore, lack of differences in *Rac1b*-deficient mice does not account for its possible role in invasion. However, knock-out of *Rac1b* did not reduce tumorigenesis either, as it was observed in the previous chapter with APC mice. Bearing in mind that almost half of the tumours from the APC RAC1B^{KO} cohort were able to overcome *Rac1b* deletion, possibly by positive selection of the non-recombinant population of cells (Figure 4.2), deletion of *Rac1b* expression was verified by qRT-PCR in both, SI and LI. As it is shown in figure 5.2E, virtually all tested mice efficiently knock-downed *Rac1b*. This might suggest that, in this context, *Rac1b* does not confer an additional advantage to tumoral cells since *Rac1b*-wild type cells did not exert a selective pressure. A possible explanation could be that jointly deletion of *Apc* and *TP53* might lead to a threshold of WNT signalling activity whereby changes in other genes such as *Rac1b* are negligible. Nonetheless, this indicates that *Rac1b* is not necessary for *Apc*- and *TP53*-driven tumorigenesis and alternative models should be considered to assess its requirement for inducing tumour invasion.



<u>Figure 5.1</u>: Validation of Rac1b expression modulation in the APC P53 model. A: qRT-PCR for *mRac1b* from SI tissue of controls (APC$^{-/-}$ P53), *Rac1b* overexpressing (APC$^{-/-}$ P53 RAC1B^{OE}) and *Rac1b* deficient (APC$^{-/-}$ P53 RAC1B^{KO}) mice. B: qRT-PCR for *hRac1b* from epithelial extractions (EE) of controls and overexpressing mice. C: Western blot for Rac1b and β-actin antibodies from EE of experimental mice. The adjacent graph presents densitometry quantification of WB bands.
T-test: **= p≤0.01, ****= p≤0.0001, n.s. = not significant, p> 0.05. N(qRT-PCR) = 5 vs 3 vs 6; N(qRT-PCR EE) = 2 vs 3; N(WB EE) = 2 vs 3 vs 3.

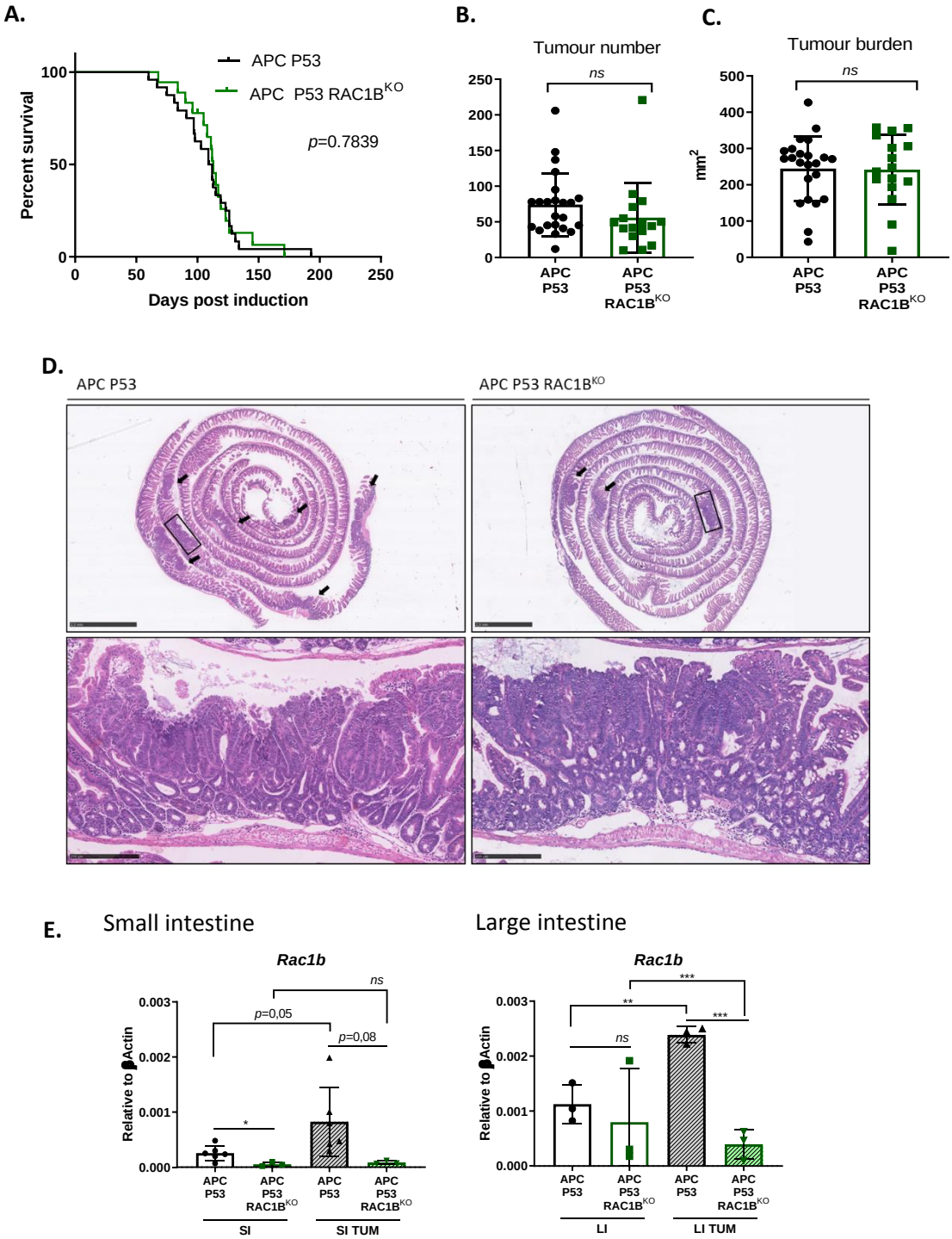


Figure 5.2: Deletion of *Rac1b* does not affect tumorigenesis in *Apc*- and *TP53*-deficient mice. A: Kaplan-Meier survival plot for APC P53 and APC P53 RAC1B^{KO} induced mice. Log-rank test for significance. B and C: tumorigenesis of cohort mice scored by the total number (B) and area (C) of tumours per mouse. D: Swiss-roll overview of SI, stained by H&E. An example of non-invasive tumour is amplified underneath. E: qRT-PCR for *Rac1b* comparing SI and LI tumours and their matching normal tissue. N(qRT-PCR SI) 6 vs 6; N(qRT-PCR LI) 3 vs 3.
Mice in the cohort (n): APC P53 = 24, APC P53 RAC1B^{KO} = 18.
T-test: **= p≤0.01, ***= p≤0.001, n.s. = not significant, p>0.05.

5.2 *Rac1b* cooperates with *TP53* and *APC* deletion to increase intestinal tumorigenesis

In the same way as with the *Rac1b*-deleted mice, validation of overexpression with *APC* and *TP53* deletion was carried out by acute induction of $APC^{-/-}$ and $APC^{-/-}$ $RAC1B^{OE}$ mice (*VillinCreER*^{T2} *Apc*$^{fl/+}$ *TP53*$^{fl/fl}$ *Rosa26*$^{+/+}$ and *VillinCreER*^{T2} *Apc*$^{fl/+}$ *TP53*$^{fl/fl}$ *Rosa26*$^{lsl\text{-}Rac1b/lsl\text{-}Rac1b}$ respectively*)*. qRT-PCR for *hRac1b* was conducted and showed a robust overexpression, as well as when Rac1b was evaluated WB (Figure 5.1). Therefore, overexpression of *Rac1b* was achieved efficiently and mice were subjected for tumour assessment.

5.2.1 *Rac1b* overexpression in an ageing cohort of APC P53 mice

Control (APC P53) and mice with *Rac1b* overexpression (APC P53 $RAC1B^{OE}$) were tamoxifen induced and closely monitored for endpoint signs. Surprisingly, the first disease symptoms in the APC P53 $RAC1B^{OE}$ group were noticeable after 55-65 days from induction, substantially earlier than the control mice which were not symptomatic of disease until about 90 days post-induction. Consequently, APC P53 $RAC1B^{OE}$ mice required to be terminated significantly earlier and presented an average survival of 79 days, compared to APC P53 mice which show an average survival of 113 days (Figure 5.3A). Despite shortened survival, mice in the APC P53 $RAC1B^{OE}$ group developed significantly more tumours than control APC P53 mice (60 vs 126 tumours respectively, p=0.0054, figure 5.3B). Mice with *Rac1b* overexpression developed lots of small non-invasive adenomas across the intestine, mainly originated in the duodenum and the jejunum and much less frequently in the ileum and colon (Figure 5.3D). Consequently, very few presented prolapses nor bleeding as signs of their performance, and pale feet and weight loss were the best references to control for their disease evolution. Both, APC P53 and APC P53 $RAC1B^{OE}$ mice were culled at the same clinical endpoint criteria since their overall tumour burden is unchanged (Figure 5.3C). Therefore, I can state that *Rac1b* overexpression together with *Apc* and *TP53* deletion significantly reduces mouse survival as a result of an increased intestinal tumour formation.

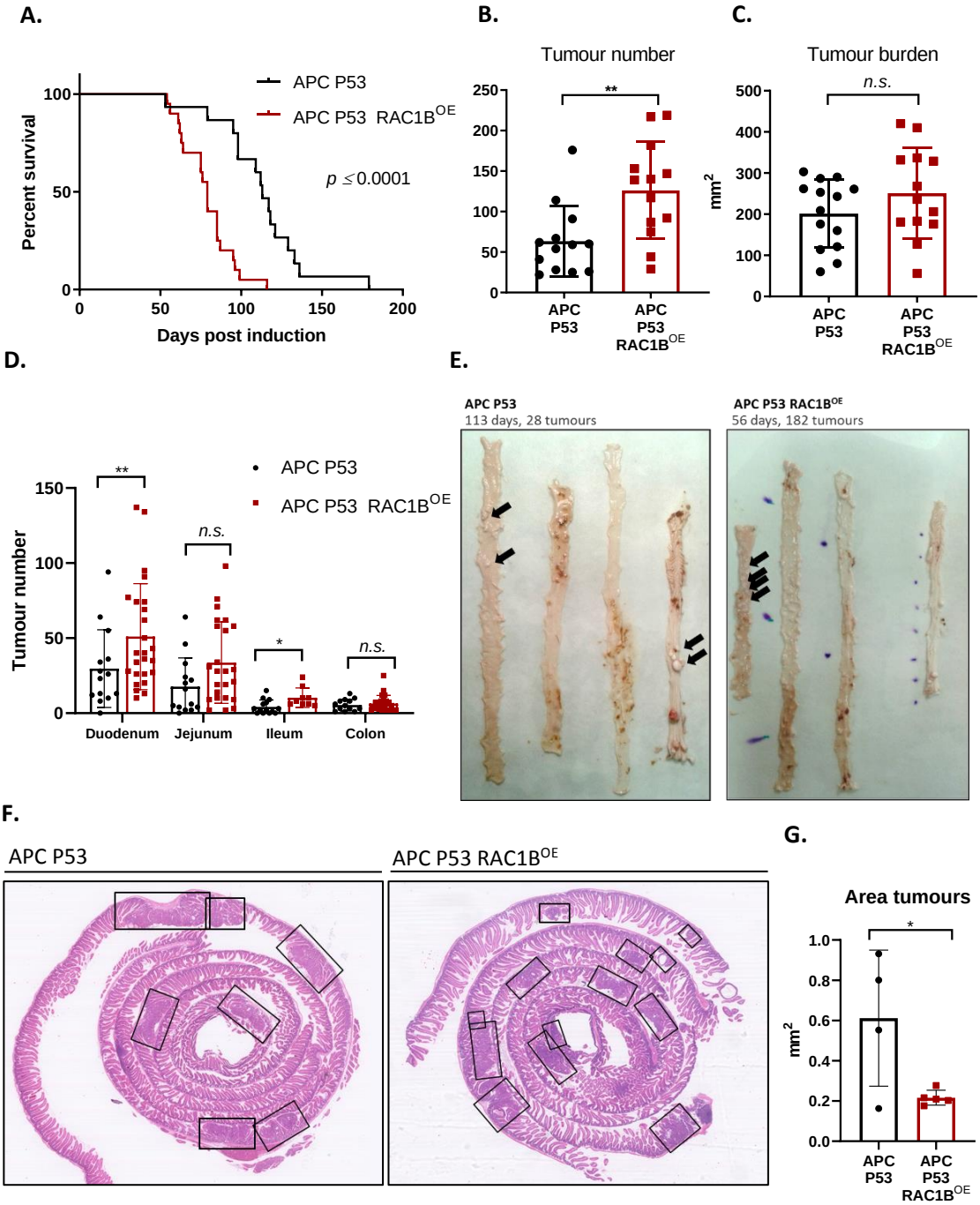
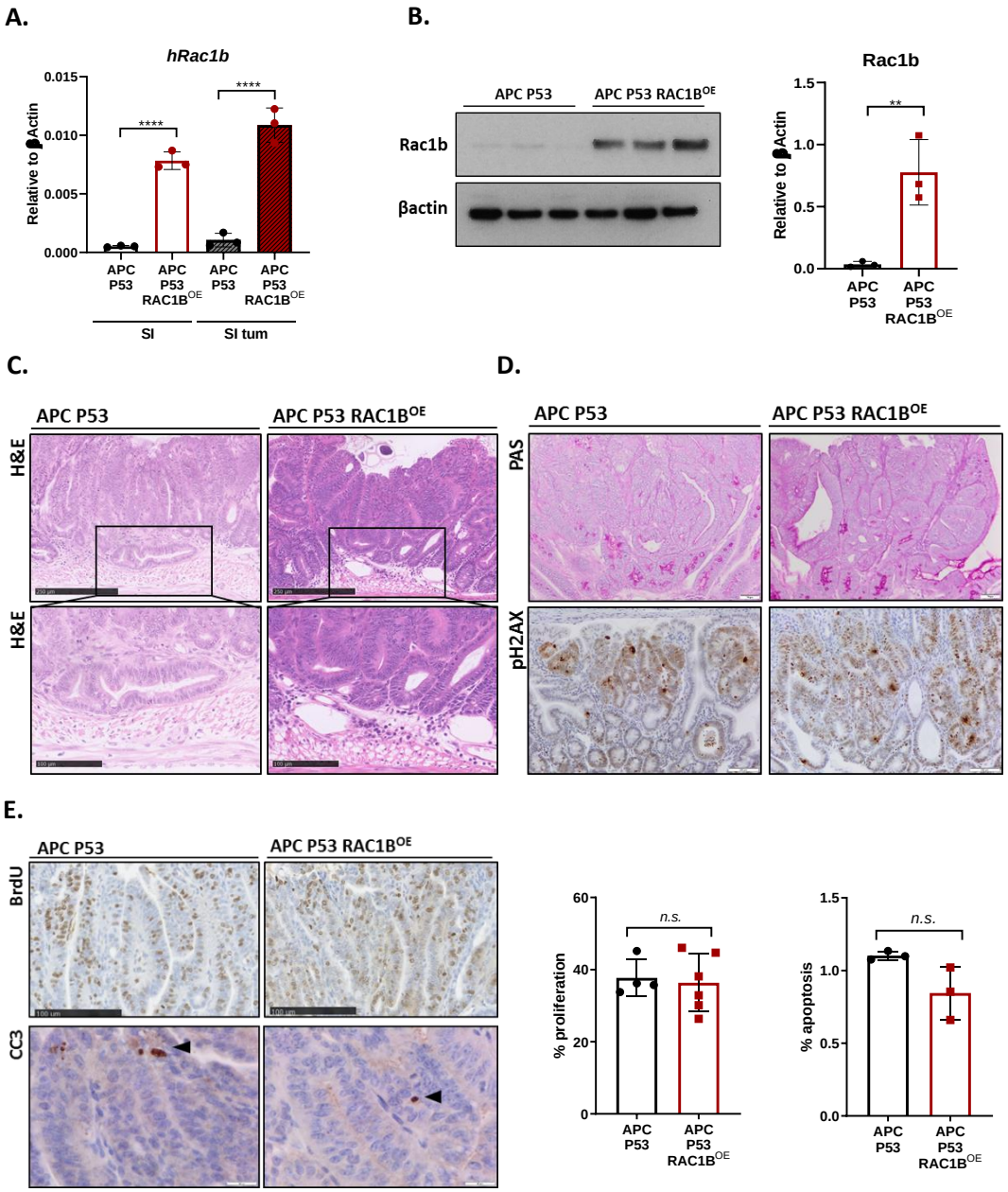


Figure 5.3: Overexpression of *RAC1B* promotes tumorigenesis and decreases survival in the APC P53 model. A: Kaplan-Meier survival graph for induced APC P53 and APC P53 RAC1B^{OE} ageing cohort mice (Log-rank test). B: Number of intestinal tumours scored when dissected. C: Total area of developed tumours. D: Number of tumours plotted based on their intestinal origin. E: Representative picture of a control and *Rac1b* overexpressing mouse intestines when dissected. Tissue was fixed on methacarn to highlight tumoral areas and some tumours are marked with arrows. C: SI overview as Swiss-rolls, stained in H&E. Squares mark tumours in the section. G: Area of tumours measured on H&E stained Swiss-rolls from 4 and 5 APC P53 and APC P53 RAC1B^{OE} mice, respectively. T-test: *= p≤0.05, **= p≤0.01, n.s.= not significant, p>0.05. Mice in the cohort (n): APC P53 = 15, APC P53 RAC1B^{OE} = 20. N(area tumours) = 4 vs 5.

5.2.1.1 Tumour histological study

Similar to the control animals used for the APC P53 RAC1B^{KO} cohort, mice in this experiment demonstrated a rapid tumour initiation phenotype and a lack of tumour invasion. Both APC P53 cohorts with either RAC1B^{KO} or RAC1B^{OE} developed comparably and only mice with *Rac1b* overexpression showed a significant difference (decreased survival and increased tumour number). In order to mechanistically understand how Rac1b drove this rapid tumour formation, a number of analyses were carried out. First, I confirmed *Rac1b* overexpression by qRT-PCR from SI normal tissue and tumours (Figure 5.4A) and WB from SI epithelial extractions (EE) (Figure 5.4B) and saw that they maintained expression of human *Rac1b* over time. I then performed some general histological analysis on the tumours for further characterisation. Analysis of H&E stained sections revealed that APC P53 mice developed larger tumours than mice from APC P53 RAC1B^{OE} group (p=0.033, Figure 5.3G). However, these were not invasive since no tumour cells reached either the submucosa or the muscularis layers (Figure 5.4C). On the contrary, APC P53 RAC1B^{OE} mice formed lots of small adenomas predominantly in the first 20cm of the intestine, with an average size of 0.22mm^2. They likely originated from the crypts and none of them broke into the submucosa either. PAS staining showed a similar pattern of mucus glycoproteins stained and both had a comparable DNA damage measured by pH2AX IHC (Figure 5.4D). Since mice were injected with a cell proliferation labelling reagent (BrdU) two hours before culling, sections were subjected to BrdU$^+$ IHC for proliferation assessment. Scoring was done manually using ImageJ and at least 15.000 tumoral cells from 12-18 tumours among different experimental mice were used for analysis. As it is shown in Figure 5.4E, there are no differences in their rate of proliferation, since both have an almost identical percentage of positive BrdU cells of 37%. Same scoring methodology was used to measure their rate of apoptosis by cleaved caspase 3 staining and again, not significant differences were observed (1.10% in APC P53 and 0.84% in APC P53 RAC1B^{OE} tumours, p=0.071, Figure 5.4E). Overall, these reaffirms the lack of invasion from this model and shows that tumours do not present major differential features.

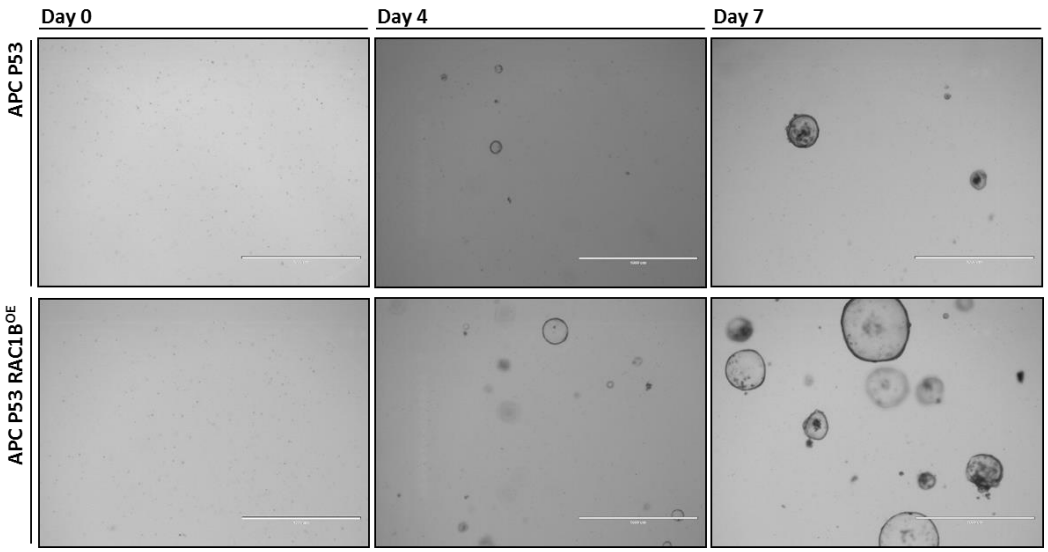


<u>Figure 5.4</u>: Endpoint tumours do not present morphological differences. A and B: Validation of human *Rac1b* overexpression in cohort mice by both qRT-PCR from SI and SI tumour tissue (A) and WB from EE (B). C: Representative H&E staining of SI tumours. Invasive front of the tumour is amplified, demonstrating the lack of invasion. D: PAS staining and pH2AX IHC to evaluate goblet cell population and DNA damage within tumours, respectively. E: Representative images of BrdU and CC3 staining. Positive and total tumour cells were manually scored. Proliferation and cell death scores are plotted for each group.
T-test: **= p≤0.01, ****= p≤0.0001, n.s. = not significant, p>0.05. N (RT-qPCR, WB and histology) = 3 vs 3, N (BrdU) = 4 vs 6, N (CC3) = 3 vs 3.

5.2.1.2 Rac1b overexpression enhances tumour clonogenicity in vitro

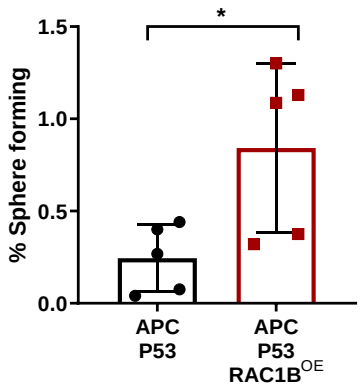
When mice were sampled after culling, tumour adenoma cells were isolated for subsequent *ex vivo* organoid culture in Matrigel. This protocol allows primary tumour organoid lines to be established directly from experimental cohort mice. I conducted clonogenicity assays on these tumour organoids lines to determine the capacity to form a clone from a single tumoral cell. Organoids were digested and sorted through FACS (Fluorescence Activated Cell Sorting) to ensure that the resulting cells from digestion will be alive, single and in the same number. The latter two characteristics are essential for this assay since total number of spheres could be biased if fragments or different number of cells are plated. Twenty-five Matrigel drops of 1.000 cells each were plated, and the number of spheres formed were measured after 4 days. As it is shown in Figure 5.5, organoids with *Rac1b* overexpression had significantly more clonogenic capacity than APC P53, around 4.5-fold change more in average (p=0.027). The table with the replicates illustrates the large variability of the technique in each repeat, although every time showed the same trend. The results given by this assay could be interpreted in two ways: on one side, it might indicate that higher levels of Rac1b make the cells more “stem-like” and more capable to grow and give rise the different intestinal cell lineages to form a clone. Alternatively, it could also suggest that Rac1b provides an advantage to the cells to overcome the harsh process of cell digestion and FACS, resulting in more resistant and surviving cells to ultimately form clones. Altogether, the assay demonstrates that overexpression of *Rac1b* enables increased tumorigenesis and higher clonogenicity in tumoral cells. Further analyses are required to mechanistically understanding these phenotypes.

A.



B.

	APC P53	APC P53 RAC1B	fold change
#1	0.4	1.13	**2.77**
#2	0.04	0.32	**7.7**
#3	0.075	0.375	**5**
#4	0.44	1.3	**2.96**
#5	0.268	1.087	**4.05**



<u>Figure 5.5</u>: Overexpression of *Rac1b* confers a major clonogenic capacity to APC P53 organoids. A: Representative organoid pictures during the clonogenicity assay at day 0, 4 and 7. B: Summary table of the five replicates, showing the percentage of sphere forming for group and the fold change of each experiment. Percentages are plotted in the adjacent graph. T-test: *p<0.05. N = 5 vs 5.

5.2.1.3 Intratumour heterogeneity hinders RNAseq analysis

In order to analyse the cohort by an unbiased approach and elucidate some of the *Rac1b* downstream target genes, RNA from three representative tumours of each cohort was sent for RNA sequencing. The results yielded 390 differentially expressed genes, 145 upregulated and 240 downregulated genes in the APC P53 RAC1B^{OE} group compared to the APC P53 (List of genes and detailed FastQC analyses in Appendix 2.3 and 2.5, respectively). Data set analysis by PANTHER gene ontology database indicated an involvement of inflammation and cell recognition processes. Moreover, most of the candidates seemed to be located in the cytoskeleton or the cell membrane (Figure 5.6A), in line with our previous RNAseq and BioID analysis from APC mouse model. Looking through the gene set list, some interesting targets were picked for further validation, such as *Cited1* (WNT target gene, positive TGFβ and ERBB2 regulator), *IGFBP2* (binding and regulation of insulin growth factor) or *Klk1* (involved in cell adhesion and extracellular cell matrix remodelling). Validation was conducted by qRT-PCR from original RNAseq samples but, surprisingly, very few were significantly different. The vast majority of genes analysed showed no differences between groups except for *Cited1* and *Klk1* (Figure 5.6B). After re-analysis of the data, we noticed that the FPKM (Fragments Per Kilobase Million) reads presented large variability between samples from the same group, which led to false positive significant results (Appendix 2.4). Yet there were some interesting targets that I did confirm, but these were not further validated when tested in a separate group of tumours (Figure 5.6C), indicating that none of these candidates are direct *Rac1b* effectors. This result, as well as the fact that virtually none of the cellular processes were not either up- or down-regulated specifically, suggests that an alternative approach is needed to determine Rac1b mechanisms of action.

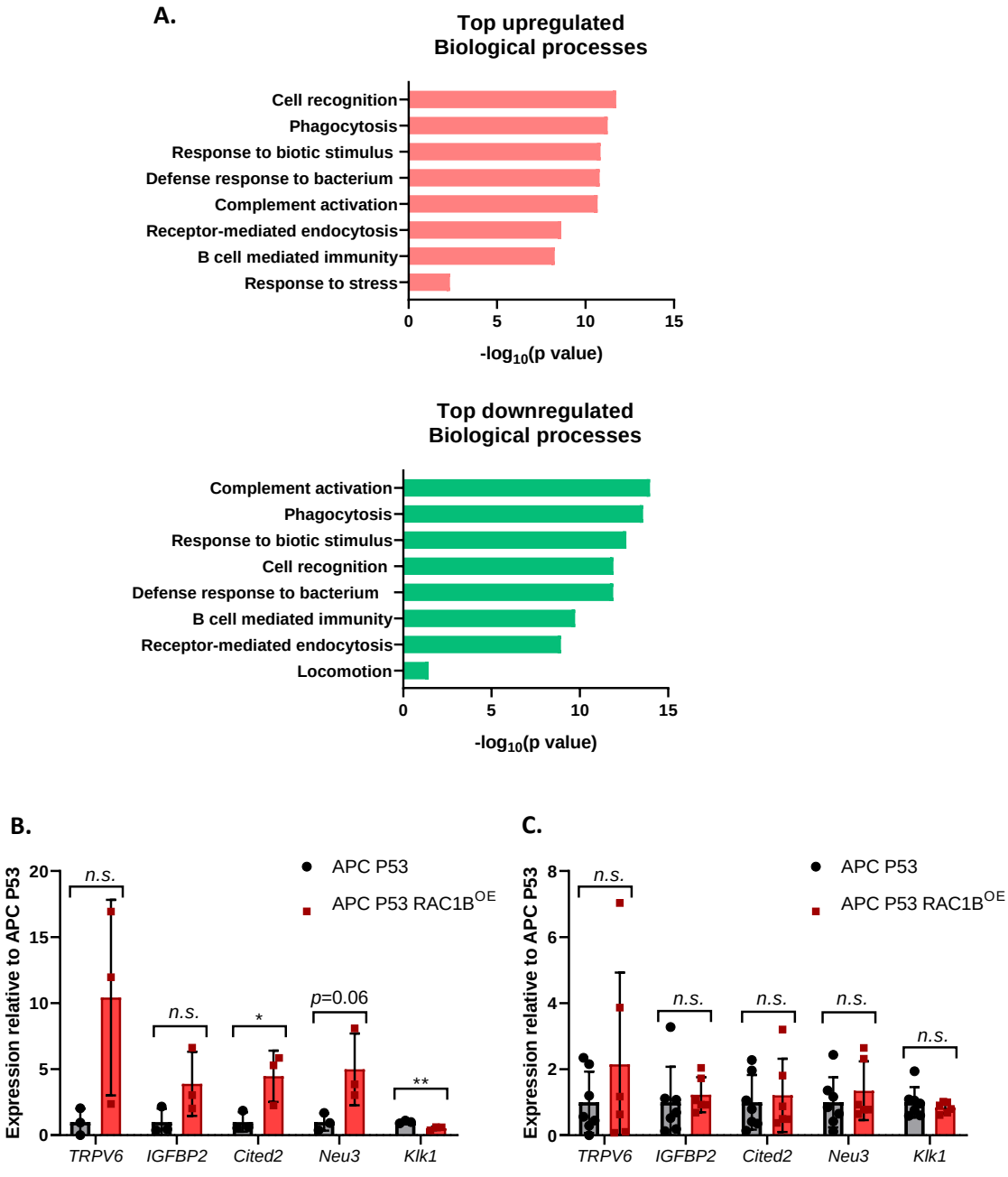


Figure 5.6: RNA sequencing analysis of endpoint tumours. A: Top up- and down-regulated biological processes from RNAseq differentially expressed genes (DEG) comparing APC P53 and APC P53 RAC1B^{OE} tumours. Gene ontology analyses were performed using Pantherdb. B: qRT-PCR validation of five DEG in the sequenced RNA samples. C: qRT-PCR validation of the same targets in a larger set of tumour samples.
T-test: *= p≤0.05, **= p≤0.01, n.s. = not significant, p>0.05. N (RNAseq) = 3 vs 3. N (qRT-PCR validation) = 7 vs 6.

5.2.2 Pre-tumorigenic model as a strategy to study tumour initiation

The endpoint tumours have shown high similarity histologically and expression-wise, not demonstrating any distinguishable characteristic between genotypes. Yet, there is an indubitable difference in the percentage of survival and tumour formation driven by *Rac1b* overexpression. Given that the model emulates an early tumour initiation stage rather than an advanced tumour invasion phenotype, analyses of the endpoint tumours might not be the most informative source and might hinder to detect early occurring events. High levels of Rac1b accelerated and enhanced tumour formation, leading to changes that, perhaps, could be more noticeable in the murine intestine before tumours originated. Following this rationale, I sought to investigate the function of Rac1b at an early stage of tumorigenesis. Therefore, I designed a pre-tumorigenic model with the same genotypes as for the aging cohort (APC P53 and APC P53 RAC1B^{OE}) but with a fixed termination day. 31 days after induction was chosen as ending point since this is approximately half of the average survival for the APC P53 RAC1B^{OE} group. I aimed that by then, some tumours would have started to form and either them or the normal intestinal tissue could provide suggestions as to the alterations driven by Rac1b. To ensure the maximum consistency of the intestinal regions analysed between mice, a dissection plan was carefully designed (Figure 5.7A). Samples were then subjected for histological and molecular analysis.

5.2.2.1 Early adenomas resemble aging cohort phenotype

When mice were dissected, intestine looked disease-free and virtually no tumours were detected macroscopically. I then proceed to conduct a histological and molecular characterization of the model. Histologically, there were no major changes by H&E (Figure 5.7C). Despite being reported that Rac1b could expand the Paneth cell population, neither Paneth nor goblet cells were affected, as the Lysozyme IHC and qRT-PCR from the corresponding markers show (Figure 5.7C and D). Crypt cells did not present higher nuclear β-catenin staining upon *Rac1b* overexpression and neither there seemed to be more DNA damage, measured by pH2AX. Besides, the WNT signalling pathway was not more activated in the APC P53 RAC1B^{OE} group, stem cells markers were equally expressed in both groups and consequently, proliferation was not increased (Figure 5.7B). However, I did notice some tumour-like small lesions in the sections. To confirm that these cells have lost the wild type

copy of *Apc*, sections were stained for β-catenin. Nuclear β-catenin staining was detected in cells within the lesion, therefore characterizing them as early adenomas or microlesions (Figure 5.8C). Moreover, quantification of the lesions from serial sections of the Swiss rolls revealed that mice with *Rac1b* overexpression developed significantly more early adenomas compared to the APC P53 group (an average of 65 and 26 microlesions respectively) (p=0.024, Figure 5.8B), indicating that the increased tumorigenesis in the aging cohort driven by Rac1b is reproducible in a pre-tumorigenic model and noticeable already just 31 days after induction. Since I did not observe a phenotype looking at the normal tissue of these mice, I decided to repeat a similar histological analysis on the microlesions. Proliferation score showed that both groups had an equivalent rate of cell growth, reaffirming again the lack of proliferative changes by Rac1b as with the endpoint tumours (34.8% for APC P53 and 34.7% for APC P53 $RAC1B^{OE}$ group, Figure 5.8D). However, when apoptosis was measured by $CC3^{+}$ IHC, microlesions from mice with *Rac1b* overexpression had significantly less cleaved caspase 3 positive cells than WT group (2.7% vs 1.6% in APC P53 and APC P53 $RAC1B^{OE}$ respectively, p=0.002, Figure 5.8D). Moreover, when apoptosis was scored by quantification of apoptotic bodies by H&E staining, a similar outcome was observed, validating the reduced cell death in these early adenomas with *Rac1b* overexpression (2.4% vs 1.3% in WT^{OE} and $RAC1B^{OE}$ respectively, p<0.0001, Figure 5.8D). Hence, this result would suggest that the increased formation of early lesions and following tumours in the APC P53 $RAC1B^{OE}$ group might be due to a diminished rate of apoptosis, which would allow transformed cells to pass the cell death check points and survive, ultimately growing as tumours.

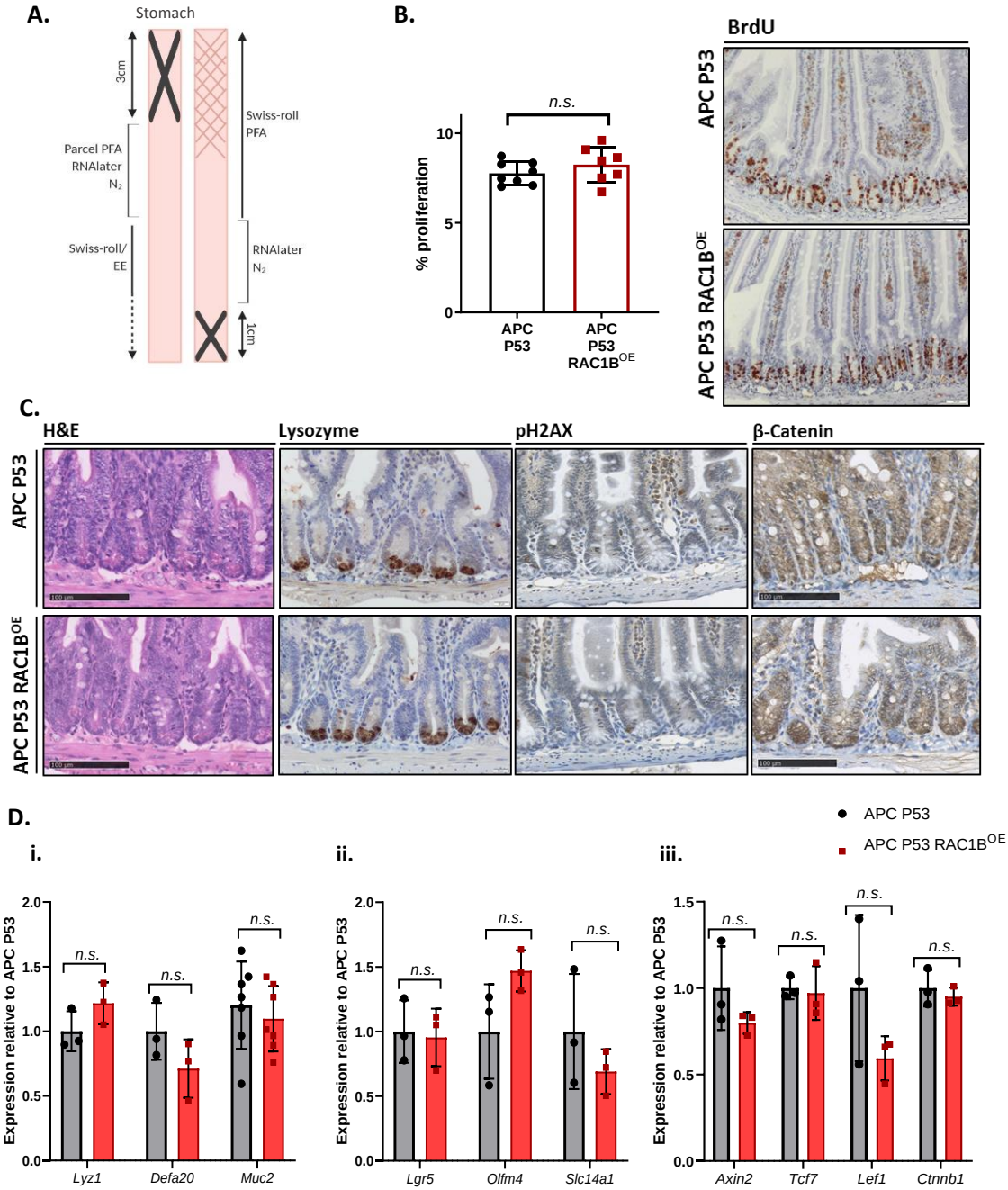


Figure 5.7: Characterisation of the day-31 pre-tumorigenic model. A: Schematic representation for dissection protocol of day-31 mice. For the SI, the first 3 cm closer to the stomach were discarded, and the following was taken to preserve in RNAlater and in liquid N_2, as well as for parcel fixation. The remaining SI tissue was either for EE or Swiss-roll fixation. For the LI, the last cm was discarded and the following was taken for RNAlater and N2 preservation and Swiss-roll fixation. B: Staining for BrdU and proliferation score of normal intestine. Positive $BrdU^+$ cells were scored in at least 50 half crypts per mouse. C: Histological characterization of normal intestine. Tissues were stained for H&E and for lysozyme (Paneth cells), pH2AX (DNA damage) and β-catenin IHC. D: qRT-PCR from normal SI tissue for Paneth (*Lyz1*, *Defa20*) and Goblet cells (*Muc2*) (i) and stem cells markers (*Lgr5*, *Olfm4*, *Slc14a1*) (ii) and WNT/β-catenin target genes (*Axin2*, *Tcf7*, Lef1 and *Ctnnb1*) (iii). T-test: n.s.= not significant, p>0.05. N (Proliferation) = 8 vs 7. N (Histology, qRT-PCR) = 3 vs 3.

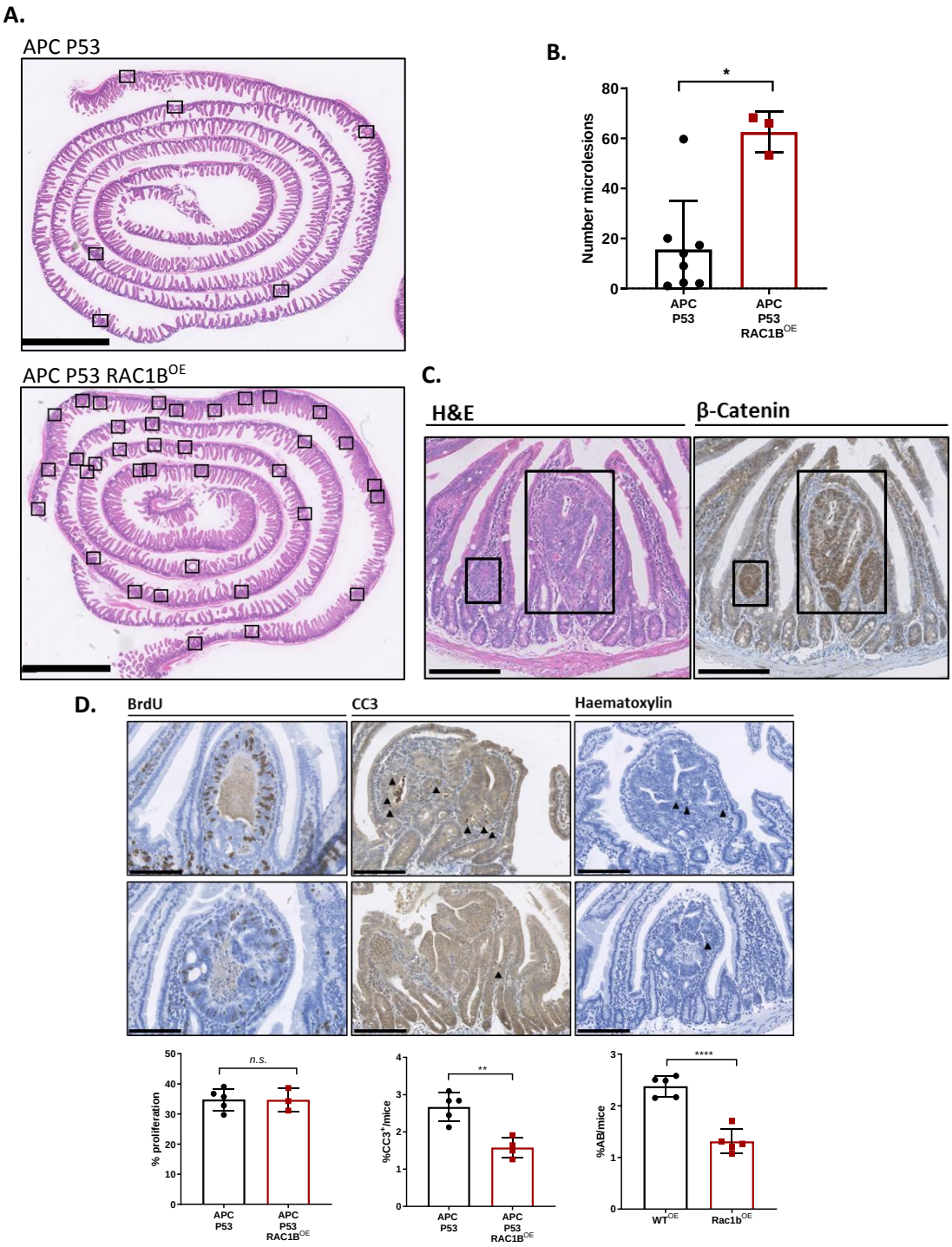


Figure 5.8: Rac1b promotes the early formation of microadenomas and reduces their apoptosis. A: Swiss roll overview of day-31 SI tissue stained by H&E. Squares marks the microlesions present in section. B: Average of microlesions detected per mouse across serial sectioning of tissue. C: Characterisation of microlesions as early adenomas by nuclear β-catenin IHC. D: Proliferation (BrdU) and cell death (CC3 and apoptotic bodies) scores of microlesions. Arrows point to apoptotic cells. Plots presents the average score of at least 5 microlesions per mouse.

T-test: *= p≤0.05, **= p≤0.01, ****= p≤0.0001, n.s. = not significant, p>0.05. N (microlesions) = 8 vs 3. N (Histology) = 5 vs 3.

5.2.3 Rac1b confers resistance to cell death through modulation of TGF-β signalling

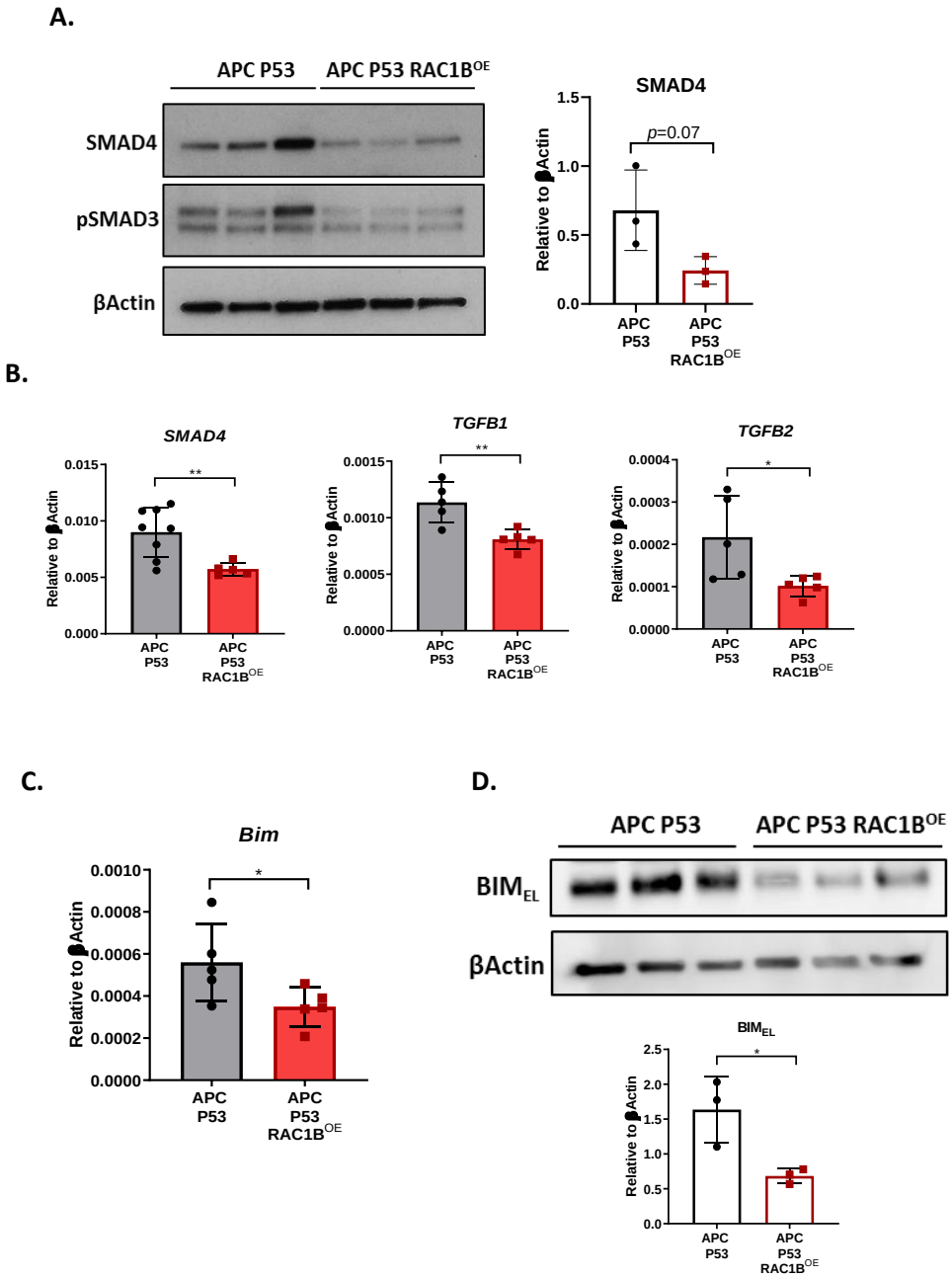
An antiapoptotic role has been suggested for Rac1b, either by inducing expression of the anti-apoptotic protein MCL1 through AKT2 activation or by stimulating NF-κB activity, which induces cell growth but inhibits cell proliferation[380,381]. Furthermore, over the last years, a link between Rac1b and the TGF-β signalling pathway has been suggested. The TGF-β signalling pathway is a superfamily of cytokines and receptors with a wide range of intracellular processes, including cell proliferation regulation, differentiation, invasion and apoptosis. In cancer, it exerts opposite roles based on the tumoral stage. At early stages, TGF-β acts as a tumour suppressor by inducing cell cycle arrest and apoptosis in tumoral cells. For example, in early colonic mouse adenomas, *Apc* mutant intestinal cells are more sensitive to TGF-β-induced apoptosis than normal epithelial cells, including *Lgr5*$^{+}$ cancer initiating stem cells[554]. During CRC progression, there is a selective pressure to diminish TGF-β signalling activity and genetic inactivation of its receptors or intracellular mediators such as SMAD4 or SMAD3 is detected in 40% to 50% of all adenocarcinomas[555]. However, high levels of TGF-β ligand associates with poor outcome in CRC patients since it facilitates metastasis and cancer progression[184], hence suggesting a pro-oncogenic role in advanced adenocarcinomas. According to Ungefroren *et al.*, Rac1b downregulates TGF-β activity by preventing SMAD3 phosphorylation, an essential step for initiation of its transcriptional program[556]. While they have mainly focussed attention on the consequences of TGF-β inhibition regarding invasion and cell migration, evaluation of its tumour suppressor function upon Rac1b deletion/overexpression has not been investigated yet. Furthermore, preliminary proteomic studies on epithelial cells from APC P53 and APC P53 RAC1B^{OE} mice suggested a deregulation in some SMAD members (data not shown). Since the phenotype observed in the early tumoral stage model is a reduction of apoptosis, which would correlate with a TGF-β inhibition phenotype, and evidence exists of a crosstalk between Rac1b and TGF-β, I decided to assess TGF-β pathway activity in the pre-tumorigenic model.

5.2.3.1 Pre-tumorigenic tissue present different TGF-β activity

qRT-PCR from day 31 mice SI tissue was carried out for the TGF-β ligands 1 and 2 and *SMAD4*. Consistently, all three markers presented significantly decreased expression in the APC P53 RAC1B^{OE} group (Figure 5.9B). SMAD4 is a common mediator for the receptor SMADs

SMAD2/3 and SMAD1/5/8, which mediates their nuclear translocation by the formation of heteromeric complexes. Lower levels of SMAD4 were also observed at the protein level by WB, and the same trend was seen when blotting against phosho-SMAD3, a canonical read out for TGF-β pathway activity (Figure 5.9A). Altogether, this would suggest that overexpression of *Rac1b* in the small intestine causes a decrease in the activity of TGF-β signalling.

One of the main mediators for TGF-β anti-apoptotic response are the Bcl2 (B-cell lymphoma 2) family proteins, which comprises anti- and pro-apoptotic members with shared homology in at least one of its four BH domains. Bim is a pro-apoptotic protein of the Bcl2 subcategory of BH-3-only proteins. Upon apoptotic stimuli, Bim binds to anti-apoptotic Bcl2 proteins and leads to MOMP-mediated apoptosis (mitochondrial outer membrane permeabilization), allowing the release of apoptotic effectors such as Bax or Bak that will lead to cleavage-dependent activation of caspases[557]. TGF-β can induce expression of Bim through p-SMAD3-dependent transcription or by ERK-dependent phosphorylation, thereby stimulating cell death. Since it was initially discovered as a direct TGF-β apoptotic target in B lymphocytes[558–561], building evidence has demonstrated Bim as an essential target for TGF-β cell death cues. Furthermore, it was shown that Bim expression mediates sensitivity of colorectal cancer tumoral cells to TGF-β-induced apoptosis[562]. In order to assess whether the reduced TGF-β activity in the pre-tumorigenic model accounts for the decrease in apoptosis, I looked at the levels of the anti-apoptotic TGF-β target gene *Bim*. qRT-PCR from day-31 murine intestine demonstrated almost twofold reduction in *Bim* transcript in the APC P53 RAC1B^{OE} mice compared to controls (Figure 5.9C). This reduction was also detected at the protein level, showing a substantial decrease of the BIM_{EL} isoform in mice with *Rac1b* overexpression, the most abundant isoform (Figure 5.9D). Collectively, these data suggest that the low apoptotic microlesions in the APC P53 RAC1B^{OE} group could potentially be a consequence of a Rac1b-induced reduction of TGF-β activity and a resultant downregulation of the pro-apoptotic gene *Bim*.



<u>Figure 5.9</u>: TGF-β signalling is decreased in *Rac1b* overexpressing mice. All analyses were made on SI epithelial extractions from pre-tumorigenic (day 31) mice. A: WB for SMAD4, pSMAD3 and β-actin as housekeeping protein. Densitometry of SMAD4 bands is plotted comparing both groups. B: qRT-PCR for *SMAD4* and *TGF-β* ligands 1 and 2. C and D presents the downregulation of Bim at the RNA (C) and protein (D) level.
T-test: *= p≤0.05, **= p≤0.01. N (WB) = 3 vs 3. N (qRT-PCR) = 8 vs 5.

5.2.3.2 Ex vivo organoid culture to study Rac1b-TGF-β modulation

In order to determine whether TGF-β-induced apoptosis is impaired by *Rac1b* overexpression, I cultured intestinal organoids from the cohort mice as three-dimensional organoids embedded in Matrigel. I assumed that to form a tumoral microlesion, cells must have lost the remaining copy of the *Apc* gene, and therefore, the genotype used for modelling these lesions in culture was *VilCreER^{T2}Apc$^{fl/fl}$TP53$^{fl/fl}$* with or without *Rac1b* overexpression (APC$^{-/-}$ P53 RAC1B^{OE} or APC$^{-/-}$ P53 respectively). Mice were induced, dissected 4 days post induction and epithelial cells extracted for either *ex vivo* organoid culture or RNA and protein characterization. As it is shown in Figure 5.10F, mice had a robust overexpression of *Rac1b* and more interestingly, I could also validate the inhibition of Bim by WB in this model. To test their TGF-β activity and the efficiency with which they respond to the signal, organoids were treated with TGF-β1 ligand for an acute stimulation. This is a commonly used experiment in many different 2D cell lines and our lab has confirmed the same procedure in organoids (data not published), with 2 hours TGF-β exposure leading to a significant induction of expression of TGF-β responsive genes. *SMAD7* is an inhibitor of TGF-β receptor 1 (TβR1) but also functions as a read out of the pathway, since its expression is increased following TGF-β treatment. RNA from organoids treated with TGF-β for 2 hours was extracted and levels of TGF-β target gene induction analysed. I found a robust induction of *SMAD7* in control APC$^{-/-}$ P53 organoids and this was significantly impaired in organoids overexpressing *Rac1b* (p=0.0013, Figure 5.10A). *p21* and *p15* are Cyclin-dependent Kinase (Cdk) inhibitors which are induced by TGF-β to halt cell cycle progression[563]. They are also early responsive genes, and their expression was significantly increased in the control group following acute TGF-β treatment (Figure 5.10A). Again, this response was significantly attenuated in organoids with *Rac1b* overexpression (p=0.011). Additionally, similar to the *in vivo* experiments, the expression induction of *Bim* was also significantly lower in organoids overexpressing *Rac1b* (p=0.024, Figure 5.10A). Furthermore, treatment of TGF-β during different time points (2, 4, 6 and 24 hours) and protein analysis confirmed the diminished activation of *Bim* seen by qRT-PCR and demonstrated that organoids with *Rac1b* overexpression never reach *Bim* induction levels observed in APC$^{-/-}$ P53 organoids. This was the case even following 24 hours treatments with TGF-β (Figure 5.10B). Accordingly, these data strongly suggest that high levels of *Rac1b* attenuates the response to TGF-β stimuli, resulting in a significant diminution in target genes that mediate cell cycle suppression and induce apoptosis.

I next sought to determine if Rac1b impacts on the physiological outputs of TGF-β treatment. In organoids with *Apc* deletion, treatment with TGF-β causes organoid death within 24 hours[554]. To determine the effect *TP53* deletion has on this phenotype, I treated $APC^{-/-}$ P53 organoids with TGF-β for an apoptotic assay. Organoids were plated as fragments, allowed to form spheres for a day and then treated with increasing concentrations of TGF-β. Organoid morphology was scored 24 hours later and classified as either dead (broken organoids with dispersed cells and debris) or alive (intact organoids without accumulation of debris)[439] (Figure 5.10E). At this time point, I observed a clear dose-dependent increase in the percentage of dead organoids following TGF-β treatment (Figure 5.10C). Interestingly, the proportion of dead organoids was lower in organoids with *Rac1b* overexpression at all TGF-β doses, with the clearest differences at 1ng/ml and 3ng/ml (p=0.02 at both doses). To confirm organoid death was occurring via induction of apoptosis, I repeated the TGF-β treatment for 24 hours exactly as before using 3ng/ml, and the Cell Event™ Caspase3/7 Green Detection Reagent from Invitrogen was added for 7 hours before analysis. This reagent has a nucleic acid binding dye conjugated to a four amino acid peptide (DEVD), which is cleaved when Caspase-3 or -7 are activated, allowing the dye to bind the DNA and releasing a fluorescent signal. Thus, organoids containing apoptotic cells emit green light. Photos at 24 hours were taken and green organoids were scored as area of green pixels relative to total organoids area in the Matrigel drop. As is shown in Figure 5.10D, in normal growth conditions organoids had equal amounts of cell death (approximately 15%). However, when TGF-β was added to the media for 24 hours, cell death increased to ~40% in the $APC^{-/-}$ P53 group but only increased slightly to ~20% in the $APC^{-/-}$ P53 $RAC1B^{OE}$ (p=0.028), recapitulating the significant decrease in organoid death induction seen before. Therefore, these assays demonstrate that overexpression of *Rac1b* in this organoid system, as in the pre-tumorigenic mice models, confers resistance to cell death which is potentially due to a downregulation of the gene *Bim* through an inhibition of the TGF-β signalling pathway.

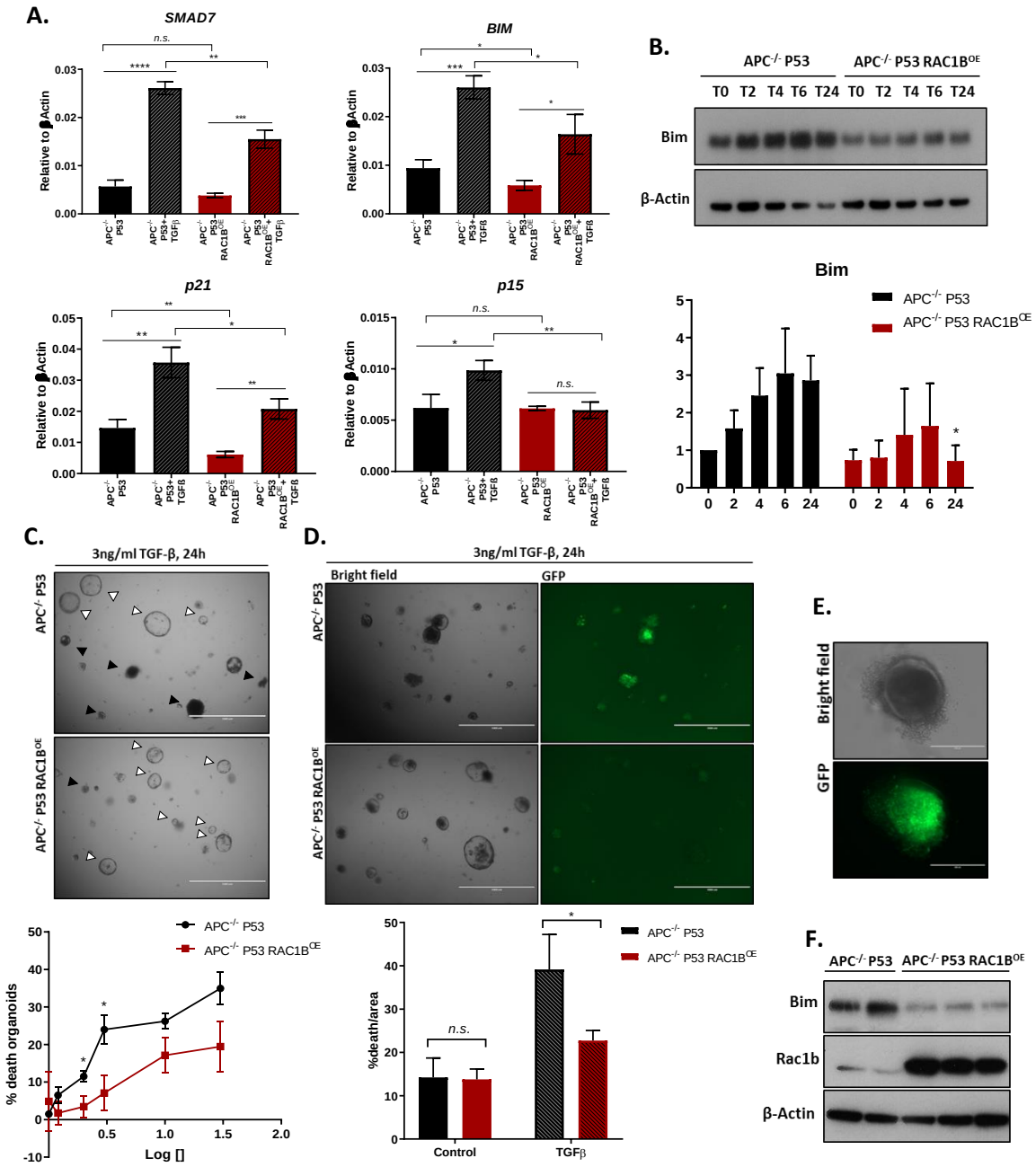


Figure 5.10: Rac1b impairs TGF-β pathway activation. A: qRT-PCR of early responsive genes (*SMAD7*, *Bim*, *p21*, *p15*) from acute (2 hours) TGF-β treatment in organoids. Total expression is the average of three technical replicates within three different cell lines per genotype. B: Response of Bim under a TGF-β time course treatment in three organoid cell lines. C: Pictures of APC$^{-/-}$ P53 and APC$^{-/-}$ P53 RAC1B^{OE} organoids after 24 hours of TGF-β treatment in the media. Graph underneath presents the percentage of cell death. D: Same conditions as in (C) but Caspase3/7 Green Detection Reagent was added to the media to allow visualisation of cleaved caspase 3. Percentage of cell death was measured by the area of green pixels relative to total organoid area. E: Representative picture of a dead organoid (bright field) and validation of the reagent (GFP), amplified at 40x. F: Validation of *Rac1b* overexpression and *Bim* downregulation in the organoid cell lines used for TGF-β treatment. Black and white arrows for death and alive organoids, respectively. T-test: *= p≤0.05, **= p≤0.01, ***= p≤0.001, ****= p≤0.0001, n.s. = not significant, p>0.05. N = 3 vs 3. N (WB) = 2 vs 3.

5.2.3.3 SMAD4 as a potential candidate for Rac1b-dependent TGF-β regulation

TGF-β is a highly mutated signalling pathway in CRC, favouring tumour initiation through inactivating mutations in most of its members. In order to assess whether expression of *RAC1B* may be associated with TGF-β signalling in human samples, an analysis of the COAD TCGA dataset was conducted. A pre-defined list of 43 genes integrating the TGF-β pathway was used to evaluate COAD tumour samples based on the presence or absence of mutations. Analysis of those samples based on *RAC1B* PSI values showed that samples with lower *RAC1B* PSI ($RAC1B^{low}$ group) are more likely to present mutations in the TGF-β pathway compared to tumours with higher *RAC1B* PSI ($RAC1B^{high}$) (Figure 5.11A, p=0.0009). Altogether, this suggests that high *RAC1B* expression may provide an alternative, non-mutational route, via tumours may overcome the tumour suppressive function of TGF-β signalling.

TGF-β pathway have two mechanisms to induce Bim expression. Upon binding of TGF-β1 ligand to receptor, sustained induction of Bim is mediated by p-SMAD3-dependent transcription of the transcription factors Runx1 and FOXO3, which will induce a delayed but sustained expression of Bim[564,565]. Parallelly, p-SMAD3 can induce the expression of *MKP2*, which will dephosphorylate and deactivate ERK1/2, therefore preventing ERK1/2-dependent degradation of Bim[560]. Interestingly, TCGA analysis of *MPK2* expression revealed that $RAC1B^{high}$ tumours express lower levels of *MKP2* compared to tumours with high *RAC1B* expression (Figure 5.11B, p≤0.0001). This shows that high levels of *RAC1B* correlate with lower expression of TGF-β target genes, suggesting its mediated downregulation, and presents a possible mechanism whereby Rac1b could downregulate Bim expression.

Even though I have associated a link between Rac1b and the cell death induced by TGF-β signalling, the mechanism through which this modulation occurs is unknown. *SMAD4* is the most mutated TGF-β pathway gene in CRC, being altered in about 13% of all CRCs[70]. Bearing in mind the initial downregulation of SMAD4 observed in the pre-tumorigenic intestinal tissue at both the RNA and protein level (Figure 5.9), I hypothesised that Rac1b might suppress TGF-β signalling by decreasing *SMAD4* expression. Interestingly, performing the same TCGA analysis looking specifically for mutations in *SMAD4* demonstrated that tumours harbouring *SMAD4* mutations express lower levels of *RAC1B* compared to $SMAD4^{WT}$ tumours (Figure 5.11C, p=0.0004). Hence, this again suggest that tumours with high *RAC1B* expression are not under as strong selective pressure to loss the tumour suppressive functions of SMAD4.

To investigate whether SMAD4 expression is altered following *Rac1b* overexpression, $APC^{-/-}$ P53 and $APC^{-/-}$ P53 $RAC1B^{OE}$ organoid protein lysates were used to analyse SMAD4 expression by WB. Consistent with my hypothesis, organoids with *Rac1b* overexpression had significantly lower expression of SMAD4 compared to controls (Figure 5.11D). Moreover, IHC staining of the endpoint tumours with a SMAD4 antibody showed a decreased immunopositivity in samples from the APC P53 $RAC1b^{OE}$ group compared to controls (p=0.01, Figure 5.11E). Interestingly, after 2 hours of TGF-β treatment, there was no downregulation of pSMAD3 levels by WB (Figure 5.11F). This suggests that phosphorylation of SMAD3 is not directly altered by Rac1b overexpression. Altogether, this data points to SMAD4 as a potential target through which Rac1b downregulates TGF-β signalling activity.

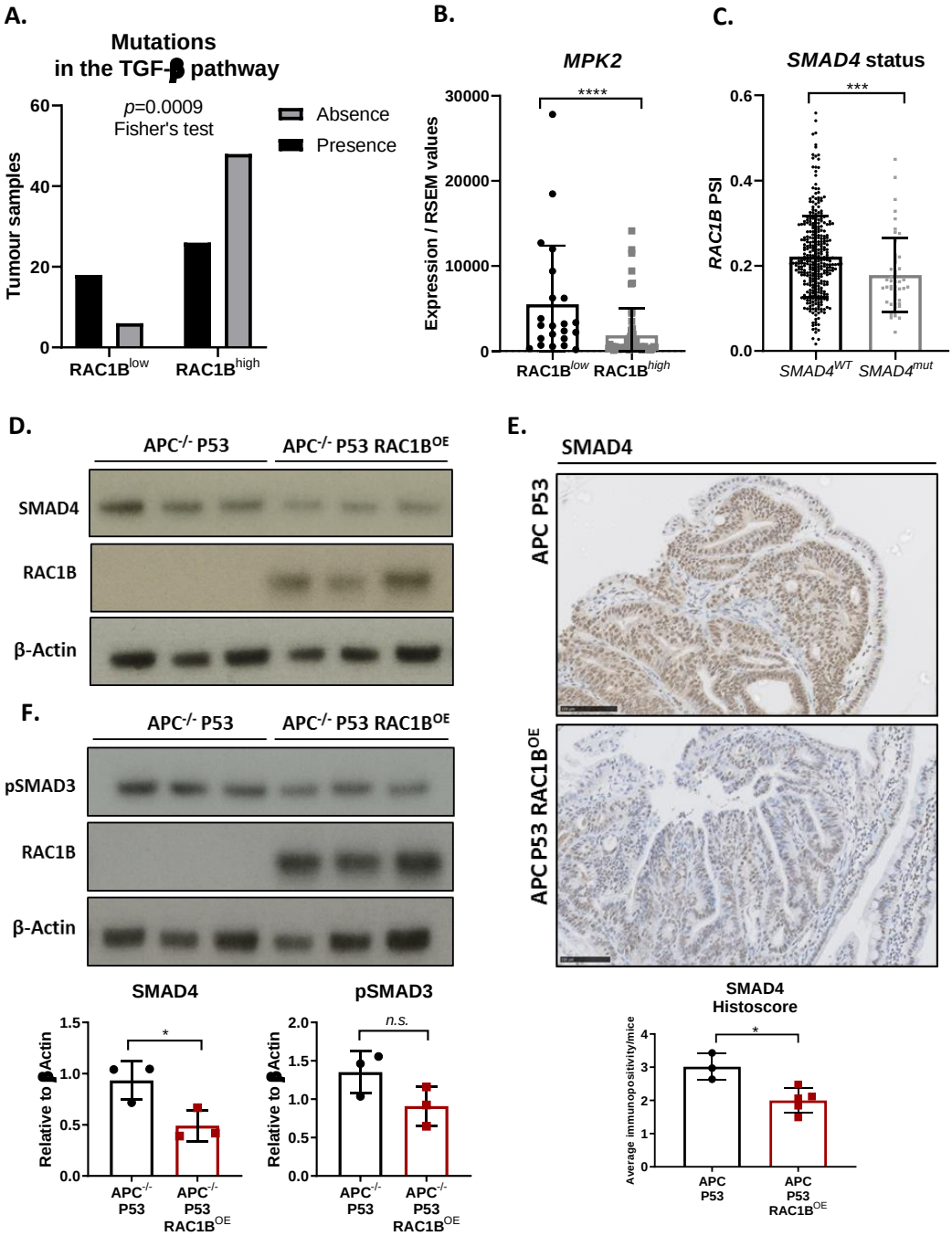


Figure 5.11: *SMAD4* as a potential Rac1b target gene. A: Evaluation of mutations in the TGF-β pathway by analysing a pre-defined set of genes from CBioPortal and associating the presence or absence of mutations in tumours to *RAC1B* PSI expression (Fisher's test). B: RSEM counts of *MKP2* comparing RAC1B^{low} and RAC1B^{high} TCGA tumours (Mann-Whitney test: ****= p<0.0001). C: Same as (A) but analysing individually the presence or absence of mutations in *SMAD4* (Mann-Whitney test: ***= p<0.001). D: Expression of SMAD4 and RAC1B in APC$^{-/-}$ and APC$^{-/-}$RAC1B^{OE} organoids evaluated by WB. E: IHC for SMAD4 of endpoint tumour cohort samples. Histoscore comparing APC P53 and APC P53 RAC1B^{OE} is presented underneath. F: WB presenting the expression of pSMAD3 of APC$^{-/-}$ and APC$^{-/-}$RAC1B^{OE} organoids treated for 2 hours with TGF-β. Densitometry quantifications of (D) and (F) WB are presented below the films. T-test: *= p≤0.05, n.s. = not significant, p>0.05. N = 3 vs 3.

5.3 Discussion

In this chapter I have outlined the effects of modulating *Rac1b* expression in a theoretically invasive colorectal cancer mouse model, driven by double deletion of *Apc* and *TP53*.

First of all, the present model did not lead to tumour invasion in any of the cohorts. Therefore, I could not test whether Rac1b is required or can promote invasion using this mouse model. Time is a limiting factor for modelling tumour invasion and the speed of the current cohort could have been an issue. An ideal mouse model would involve a low tumorigenic rate but prolonged survival, allowing tumours to develop and to further progress to aggressive carcinomas. Hence, alternative invasive models for studying a loss- and gain-of-function effect of *Rac1b* might be more useful and informative than the current one, such an orthotopic transplantation of colonic genetic modified organoids or the use of tail vein metastatic models.

The APC P53 model presented an average of 67 tumours and a median survival of 113 days. Deletion of *Rac1b* did not prolong survival in this model, as it did when it was depleted with *Apc* alone. p53 is an inhibitor of the WNT signalling pathway and its acute loss along with $Acp^{fl/fl}$ enhances the transcription of WNT target genes, hence increasing WNT signalling activity[566]. Therefore, it is possible that APC P53 tumours could have reached a threshold of WNT activity that deletion of *Rac1b* is not sufficient to attenuate tumorigenesis in this context. Indeed, mice with *Rac1b* overexpression did not enhance WNT signalling in this context, indicating that Rac1b might not directly modulate the pathway. In fact, this is in line with the BioID interactome analysis, which indicated that Rac1b did not interact with any core member of the pathway. Hence, deletion of *Rac1b* in the context of *Apc* and *P53* deficiency does not reduces tumorigenesis. However, it would be interesting to test if APC P53 $RAC1B^{KO}$ mice might confer an enhanced anti-tumorigenic effect if were co-treated with an inhibitor of the WNT pathway, EGFR pathway or a chemotherapeutic agent.

On the contrary, overexpression of *Rac1b* in co-operation with *Apc* and *TP53* depletion has revealed a very interesting and unknown phenotype: increased tumorigenesis compared to the control group and a consequent reduction in survival time after tamoxifen induction. Moreover, isolated tumoral cells showed that APC P53 $RAC1B^{OE}$ had an increased clonogenic capacity. However, morphological analysis and unbiased analysis by RNAseq of the endpoint

tumours did not provide much information about Rac1b mechanisms of actions, and I therefore sought an alternative approach to understand the mechanism behind these findings. A pre-tumorigenic model with a fixed endpoint was designed for the analysis of normal tissue or aberrant crypt foci, aiming to detect at early time points, the intestinal alterations driven by Rac1b overexpression. These mice presented early adenomas or microlesions which recapitulated the same phenotype as the aging cohort mice, since the APC P53 RAC1B^{OE} group had significant more in number compared to controls. Morphological analysis by two different means demonstrated that these lesions were less apoptotic. Even though the endpoint tumours were not less apoptotic, I did observe that APC P53 RAC1B^{OE} organoids performed better in the clonogenicity assay, which is also an indication of a better resistance to the harsh process of single cell digestion and FACS and a consequent improved survival.

Due to its anti-apoptotic role while it acts as a tumour suppressor in early tumours, and its previous suggested crosstalk with Rac1b, I tested the activity of the TGF-β pathway in the pre-tumorigenic model. qRT-PCR of TGF-β ligands and *SMAD4* showed reduced expression in the SI of APC P53 RAC1B^{OE} mice. Besides, a pro-apoptotic TGF-β target gene which has been demonstrated to be a mediator of apoptosis in *Apc* mutant intestinal cells, *Bim*, was also significantly downregulated. Crypt extraction and 3D organoids culture resembled the pre-tumorigenic lesions pattern of expression. For functional analysis of the TGF-β pathway, organoids were treated with TGF-β1 ligand. Organoids with high *Rac1b* were less responsive to this treatment and never reached same levels of pathway activation as the controls, even after 24 hours of treatment. Other than transcriptionally, I also showed phenotypically less TGF-β-induced apoptosis in the APC$^{-/-}$ P53 RAC1B^{OE} organoids by measuring death and caspase-3/7 assay.

Interestingly, TCGA analysis of the colorectal tumours data set revealed a strong negative association between *RAC1B* expression and the presence of mutations within the pathway. TCGA tumours that presented at least one genetic alteration among the 43 gene list pre-defined by CbioPortal expressed significantly lower levels of *RAC1B*, measured by its PSI. Therefore, this leads to hypothesise that overexpression of *RAC1B* may be an alternative pathway to inactivate TGF-β signalling in colorectal tumours.

Studies from Ungefroren *et al.* in PDAC cell lines showed that Rac1b downregulated TGF-β through blockage of SMAD2/3 phosphorylation, therefore preventing the complex

translocating to the nucleus with SMAD4. While preparing this thesis, they have further suggested that Rac1b might interact with the G-protein coupled receptor PAR2 and interfere with activation of the TβRI/ALK5, hence translating into a decreased activity of the pathway[404]. However, my experiments suggested SMAD4 is a possible candidate to mediate Rac1b-dependent regulation. Further, TCGA analysis revealed that tumours with inactivating mutations in *SMAD4* are less prone to induce *RAC1* alternative splicing. My previous transcriptomic analysis in pre-tumorigenic intestine tissue demonstrated a significant downregulation of *SMAD4* mRNA transcript and protein in mice with *Rac1b* overexpression. Further evaluation showed, on one side, that organoids with *Rac1b* overexpression (which presented a TGF-β-induced resistance to cell death) had a decreased expression of SMAD4 protein. On the other side, histological analysis of endpoint tumours demonstrated lower $SMAD4^{+}$ intensity staining in tumours with *Rac1b* overexpression. Altogether, these analyses strongly point to SMAD4 as a downstream mediator of Rac1b to negatively modulate TGF-β activity.

In summary, I suggest that Rac1b enhances cell transformation by conferring resistance to cell death in early transformed tumoral cells, likely via attenuation of TGF-β-signalling.

5.4 Future work

Even though the anti-apoptotic role for Rac1b and the involvement of TGF-β seems quite plausible, there are some caveats and missing experiments which, due to time, were not carried out and would complete this mechanistic hypothesis. On the one hand, since *SMAD4* is my candidate gene to connect Rac1b and the TGF-β pathway, it would be informative to determine whether a physical interaction exists between Rac1b and SMAD4, by either immunoprecipitation or co-immunofluorescence. On the other hand, I would like to attempt to phenocopy the resistance to TGF-β-induced apoptosis in the organoids by downregulating either *Bim* or *SMAD4* in the $APC^{-/-}$ P53 organoids, which would reaffirm these two genes as mediators of Rac1b-death resistance. Conversely, overexpression of Bim in the $APC^{-/-}$ P53 $RAC1B^{OE}$ organoids would allow us to determine whether Bim regulation mediates the observed phenotype. Likewise, it would be interesting to treat $APC^{-/-}$ P53 organoids with

morpholino to downregulate *Rac1b* and check whether apoptosis or TGF-β-apoptotic related genes increase their expression.

Interestingly, TCGA data suggest a possible correlation between *MKP2* and *RAC1B* expression in tumours. Given that MAPK/ERK pathway constitutes the non-canonical TGF-β signalling pathway, it would be interesting to validate whether Rac1b-dependent Bim downregulation occurs via downregulation of TGF-β/SMAD signalling or via ERK1/2 signalling activation. Moreover, ERK1/2 controls SMADs signalling through inhibition of p-SMAD3, suggesting another possible mechanism through which SMADs signalling could be downregulated. Inhibition of ERK1/2 by either treating the organoids with ERKi or by CRISPR followed by treatment with TGF-β would potentially reveal which pathway is essential for Bim inhibition.

Besides, an essential point to cover was whether Rac1b-driven resistance to apoptosis is specific to TGF-β or it is a general trait to any other cell death stimuli. Other people from the lab addressed this question by treating $APC^{-/-}$ P53 and $APC^{-/-}$ P53 $RAC1b^{OE}$ organoids with oxaliplatin. Survival score by Resazurin cell viability assay revealed no differences between genotypes, indicating that, at least in this model, resistance to cell death is a specific trait induced by TGF-β signalling.

Finally, I would like to screen Rac1b function in invasive human CRC organoids and test whether it maintains its inhibitory role among TGF-β or it switches along TGF-β to cooperate with its oncogenic activities.

These experiments would help to complete the suggested anti-apoptotic Rac1b role and elucidate whether Rac1b might be a good target to tackle therapeutically CRC at another level, without disturbing such a multifunctional and double-edge sword pathway as TGF- β.

Chapter 6: Discussion

6.1 Summary

In this thesis, I have described Rac1b functional roles in both normal and transformed murine intestine, and presented evidence for Rac1b tumour-related functions in human cancers, with special focus on CRC.

Initially, epithelial fractionation and RNA *in situ* experiments allowed the quantification and visualisation of *Rac1b* transcript in the intestine. I identified a crypt base preferential location for *Rac1b*, which gradually decreases towards the topmost villus region. In physiological conditions *Rac1b* is expressed at low levels and its expression increases and expands across the crypt-villus axis upon *Apc* loss. This does not necessarily define *Rac1b* splicing as a WNT target gene, as many other factors in conjunction with *Apc* are dysregulated. However, it indicates a potential role for Rac1b in processes leading to hyperplastic conditions, such as cancer. TCGA data analyses showed that seven out of twenty tumour types expressed significantly higher *RAC1B* PSI compared to matched normal samples (including CHOL, COAD, LIHC, LUAD, LUSC, READ, STAD and THCA). Interestingly, CRC is one of the cancers with the highest fold increase. Further analyses within the COAD data set revealed that *RAC1B* expression associates with poor clinical outcome, being more highly expressed in advanced and invasive tumours and correlating with lower disease-free survival and overall survival. Moreover, its expression defines a subset of Wnt activated CRC tumours within the CMS stratification.

Functional analyses demonstrated that deletion of *Rac1b* in an *Apc*-driven adenoma mouse model improves survival and decreases intestinal tumorigenesis. Potential molecular mechanisms include modulation of EGFR signalling and downregulation of the WNT signalling pathway, either through EGFR-dependent or -independent processes. Importantly, *Rac1b* depletion sensitises both mouse and human organoids to EGFRi treatment, pointing to a therapeutic benefit for *Rac1b* deletion. On the contrary, its overexpression in the same mouse model enhances tumorigenesis but is restricted to the large intestine. This phenotype did not lead to survival changes and the underlying mechanisms are unknown.

In a different mutational context, overexpression of *Rac1b* with *Apc* and *TP53* deficiency leads to an increased tumorigenesis and significant reduced survival. Mice bearing

Rac1b overexpression developed a rapid tumour initiation, potentially driven by a downregulation of TGF-β-dependent induction of apoptosis. Downregulation of *SMAD4* and *Bim* stand as potential Rac1b target candidates, the latter being also a mediator of EGFRi-induced apoptosis.

Altogether, these data suggest *Rac1b* as a pro-tumorigenic gene and provide evidence for a beneficial clinical outcome if deleted. These results are discussed in further detail below.

6.2 Rac1b is not required for intestinal homeostasis

When *Rac1b* was acutely deleted in the murine intestine, no phenotypic changes were observed, indicating that its expression is not required during physiological conditions. This is not entirely unexpected, given the low basal expression of *Rac1b*. Therefore, it would suggest that environmental or genetic events are necessary to promote its expression and to drive its functional roles. In contrast, Li *et al.* showed that deletion of *Rac1b* in normal rat colonic mucosa decreases expression of the anti-apoptotic protein MCL1 through AKT2 activation[380]. MCL1 expression was not assessed in my model and given that apoptosis in normal cells is virtually undetectable, I cannot verify whether Rac1b-AKT2-MCL1 activation also occurs.

When *Rac1b* was overexpressed, no gross phenotypic changes were observed either. This is contrary to Kotelevets *et al.* observation, who detected an increased in proliferation and WNT activity using the same model[409]. Some differences might account for it, such as the duration of the experiment or the mouse genetic background. Nonetheless, high overexpression of *Rac1b* on its own is insufficient to initiate tumorigenesis[408,409], as I also observed after 1-year overexpression of *Rac1b* with *p53* deletion (data not shown). The levels of overexpression acquired with this model are comparable to the expression of *Rac1b* in advanced tumours and such levels of expression will likely not be achieved in physiological conditions. Therefore, the lack of effect in both of these models reinforces the requirement of other oncogenic drivers to translate into a functional impact.

6.3 Rac1b as a promoter of tumorigenesis

The principal outcome from this thesis is the pro-tumorigenic role of Rac1b, which is validated in two different mouse models and in a large set of human expression data. Interestingly, these models have led to distinct molecular mechanisms underlying Rac1b function (Figure 6).

Loss of *Apc* is an initiating event in CRC and occurs in about 85% of all CRCs. When *Rac1b* is deleted in a benign adenoma model driven by *Apc* deficiency, it decreases tumorigenesis and significantly prolongs mice survival. Interestingly, about half of the knock-out tumours did not show *Rac1b* deletion, probably due to positive selection of cells lacking efficient Cre deletion of the Rac1b allele. These observations strongly indicate that tumoral cells benefit from increased *Rac1b* expression. Indeed, BaseScope and qRT-PCR experiments show that *Rac1b* expression is significantly increased upon deletion of *Apc* compared to healthy tissue, suggesting a deregulation of the alternative splicing machinery during *Apc*-driven cellular transformation (Figure 6, number 1). The lab has shown that a large proportion of splicing factors are deregulated following *Apc* deletion in the murine intestine (unpublished data). Among them is SRSF1, which has previously been identified as a key positive regulator of Rac1b AS. However, our own preliminary studies differ from previous studies. On one side, Jordan and colleagues proposed a model where SRSF1 promotes exon 3b inclusion while SRSF3 prevents it[412]. Likewise, upstream activation of signalling pathways controls the expression of splicing factors. An example for this is WNT signalling which drives SRSF3 expression therefore indicating high WNT activity inhibits *Rac1b* expression[567]. On the other side, further experiments showed that *Rac1b* expression correlated with *B-RAF*^{V600E}-driven tumours[568]. However, in other experiments carried out in the lab deletion of *SRSF1* or *SRSF3* in the murine intestine had no effect on *Rac1b* expression (data not shown). Moreover, our observations that *Apc* loss leads to an increased *Rac1b* expression further suggests that the WNT signalling in fact positively controls Rac1b AS. Apart from *in vivo* mouse experiments, whole-exome sequencing data from the TCGA database supports my hypothesis. Expression analyses demonstrate a strong positive correlation between *RAC1B* expression and WNT activated CRC and a strong negative correlation with *B-RAF* mutated tumours. Furthermore, MSI status, which along with *B-RAF* mutation constitute the driving traits for tumours developed through the serrated route, is also negatively associated with *RAC1B*. Stratification of CRC tumours according to the consensus molecular subtypes (CMS)

classification showed that *RAC1B* and the CMS1 (MSI, *B-RAF*^{mut} rich, immunologic) are negatively correlated. Therefore, previous assumptions about Rac1b regulation based on cell models appear to be not completely accurate and warrant further *in vivo* validation. Studying the splicing factors that lead to *Rac1b* expression is out of the scope of the current project, but their investigation will provide interesting insights into both *Rac1b* and other possible cancer-related alternatively spliced isoforms.

Tumours developed from APC RAC1b^{KO} mice showed decreased proliferation and downregulation of WNT signalling, identified by RNAseq and validated by qRT-PCR. Modulation of the WNT pathway by Rac1b has also been suggested in the literature, especially from Bapat's lab. They suggest that Rac1b can localise to the nucleus and promote the transcription of β-catenin target genes[386,387]. Its nuclear location could be plausible given the NLS sequence within the PBR region of Rac1b, which targets both its membrane and nuclear location[264]. However, modulation of the WNT pathway upstream of β-catenin is something we think unlikely for a number of reasons. Firstly, nuclear staining of β-catenin does not decrease upon *Rac1b* deletion. Secondly, the GTP-bound state would be expected to cause Rac1b to localise near to the membrane, where small GTPases are activated. Finally, my BioID protein interactome experiment, on top of suggesting a membrane preferential location of Rac1b, did not present any core member of the WNT pathway as a Rac1b interacting partner. Some resultant proteins might influence its signalling, such as Cyclin Y, which phosphorylates LRP6 and primes the receptor for activation upon Wnt ligand binding[468]. The possible regulation of Wnt signalling by Rac1b via Cyclin Y warrants further investigation. The NLS might be a conserved motif from other Ras small GTPases such as the Ran family, with an active nuclear function. Rac1 has a nuclear location with an important role in cell cycle progression, but further investigations are needed for the potential nuclear role of Rac1b.

The positive correlation between *RAC1B* expression and CMS2 (canonical, WNT driven) reinforced the WNT signature observed in the APC RAC1b^{KO} mice. However, its modulation may be an indirect rather than a direct effect, according to the BioID interactome. ERBB1 and ERBB2 were identified as interacting proteins and due to the enrichment of several membrane bound RTKs became the focus for further analysis. These RTKs activate diverse kinase signalling cascades such as MAPK and PI3K signalling and are frequently overactivated and mutated in cancer. Functional studies revealed that downregulation of *Rac1b* attenuates

the activation of EGFR downstream signalling, evaluated by phosphorylation of EGFR, ERK1/2 and AKT (Figure 6, number 3). Importantly, the strongest difference in activation was observed after 10 minutes upon ligand stimulation, when receptor-ligand dimers are potentially internalised at the early endosomes and either enter into the lysosomal degradation route or are recycled back to the plasma membrane. Signalling pathway modulation could be achieved at the receptor level, whereby shifting their internalisation fate from the degradation to the recycle route would increase the quantity and availability of free receptors at the plasma membrane. Related literature suggests a role for Rac1b in driving such modulation. On one side, Rac1 mediates vesicle endocytosis through its effector synaptojanin 2, whose overexpression and copy number gains promotes EGFR recycling to the cell surface and associates with metastatic breast cancer[569,570]. Moreover, other Rho GTPases mediate EGFR internalisation, such as RhoB, whose location at the late endosomes reduces receptor degradation[247]. Indeed, its overexpression in lung cancer associates with EGFRi resistance, whereby RhoB sustains phosphorylated EGFR at the plasma membrane and activates AKT signalling[571]. On the other side, several Rac1b interacting proteins are involved in both Rac1-related actin dynamics and in EGFR signalling regulation. One of the caveats of BioID is that some of the resultant proteins may not be direct interactors but neighbour proteins. However, this also provides a more comprehensive interactome landscape and identifies protein networks rather than single interactions. An interesting protein is WASF2, which mediates actin polymerisation through activation of the Arp2/3 complex and promotes invasion in breast cancer cells through constitutive EGFR endosomal recycling[572]. Likewise, Asap1 also supports this hypothesis. Several studies have suggested its pro-tumorigenic role based on enhanced EGFR recycling[475,528,532] and one of its effectors, Arf6, has already been shown to interact with Rac1b and promote endosomal recycling in neurons[535]. The abundance of membrane proteins, as well as actin-related proteins in the BioID dataset, might indicate that both Rac1 and Rac1b could localise in the vicinity of EGFR and ERBB2, its main heterodimer partner, and promote their activity by facilitating their recycling dynamics (Figure 6, number 4). However, Rac1b might perform this process more efficiently due to its levels and GEF-independent activation. The phenotype observed was a delayed and lower EGFR signalling response, pointing to a loss of efficiency in this process upon *Rac1b* loss. Targeting of RTK degradation route is a current approach for attenuating aberrantly activated signalling pathways. Therefore, characterising whether Rac1b is involved during receptor endocytosis adds a new potentially druggable target to modify this process.

Furthermore, downregulation of EGFR signalling could also account for the decreased activation of the WNT pathway. Given their cooperative role in promoting CRC tumorigenesis, mutations in the WNT and Ras pathways are among the most prevalent events. Both pathways can be simultaneously activated upon WNT ligand binding, given that the β-catenin destruction complex also promotes phosphorylation and degradation of Ras[478]. A positive feedback loop might exist from ERK, whose activation stabilises β-catenin in the cytoplasm[573,574]. Moreover, a recent work has shown that EGFR is required for FZD9B phosphorylation, an essential step to mediate internalization of WNT receptor and to conduct signal transduction in hematopoietic cells[479]. Altogether, this indicates that due to the cross-activity of both pathways, downregulation of one will likely interfere into the others activity. I cannot discard that Rac1b directly modulates WNT signalling (Figure 6, number 2), but the interactome data rather points to an indirect consequence following the modulation of other pathways such as EGFR.

Signalling output seems to be largely context dependent. Whilst co-deletion of *Apc* and *Rac1b* led to this interesting phenotype, no effect was observed when it was deleted in an *Apc*- and *TP53*-deletion model. One of the hypotheses includes the levels of WNT activity. Acute induction of this model showed that *TP53* induced a subtle but significant increase in the transcription of WNT target genes[566]. Besides, it has recently been demonstrated that p53 represses the transcription of some WNT ligands (WNT1, WNT6 and WNT7A) and its loss induces their cell-autonomous secretion in cancer cells, leading to an enhanced WNT activity[575]. This could suggest that APC P53 RAC1b^{KO} tumoral cells might reach a threshold of WNT activation where deletion of *Rac1b* alone is not enough to counteract it. A suggestive event is the cellular response upon knocking-out *Rac1b*: whereas almost half of the tumours analysed of the APC RAC1b^{KO} cohort overcame *Rac1b* deletion, all APC P53 RAC1b^{KO} tumours maintained its knock-out throughout tumorigenesis. This indicates that, in this context, expression of *Rac1b* does not exert a positive selective pressure. Although mice did not present a different phenotype, subtle molecular changes might have occurred. Further characterization could be interesting, especially to explore a potential synergism of *Rac1b* deletion under therapeutic treatments. However, a large part of the project was invested in describing other co-occurring mouse models, such as the APC P53 with *Rac1b* overexpression.

Overexpression of *Rac1b* in this same model substantially decreased survival as a result of increased tumorigenesis. These mice modelled a tumour initiation phenotype whose enhanced tumorigenesis was validated in two different cohorts: an ageing and a pre-tumorigenic time point cohort. Interestingly, the latter developed small micro lesions which nuclear β-catenin staining characterised as early adenomas. Histological analysis revealed a decreased rate of apoptosis in the APC P53 RAC1b^{OE} group, which likely allowed the progression of transformed cells and their evolution towards adenomas. Even though the endpoint tumours did not present an anti-apoptotic phenotype, clonogenicity experiments from isolated tumoral cells demonstrated a higher clonogenic rate, which also indicates that these tumours were more resistant to the stressful conditions of this assay than controls. Mechanistically, dysregulation of TGF-β signalling appeared as a feasible explanation for a number of reasons. Initially, a previous MS analysis from intestinal epithelial fractions of pre-tumorigenic mice pointed to some SMADs members being downregulated (data not shown). Moreover, TCGA analyses demonstrated that tumours with high *RAC1B* expression bear fewer inactivating mutations in the TGF-β pathway, indicating that these tumours might utilise *RAC1B* to downregulate the pathway. Furthermore, consistent studies from Ungefroren and colleagues have associated *RAC1B* expression and lower TGF-β activity[403,550]. Expression analysis in the pre-tumorigenic tissue showed a downregulation in some TGF-β core members, such as SMAD4, its ligands TGF-β1 and 2 and importantly, a decrease in its pro-apoptotic target Bim (Figure 6, number 5). Moreover, acutely deleted APC$^{-/-}$ P53 RAC1b^{OE} organoids recapitulated this phenotype and were more resistant to cell death upon TGF-β stimulation. Interestingly, resistance to cell death was a specific response from TGF-β, since treatment with oxaliplatin did not show any difference. Expression and histologic analyses point to SMAD4 as the candidate gene through which Rac1b might downregulate the pathway, but direct evidence of this is lacking (Figure 6, number 6). Although the molecular mechanism through which Rac1b modulates the pathway is unknown, decreased expression of Bim was consistently observed when Rac1b was highly expressed. Indeed, stimulation of TGF-β for 24 hours showed that Bim never reached the same expression levels as the control organoids. This is of relevance, especially due to the EGFR modulation observed in the previous cohort. Inhibitors of EGFR mediate cell death through expression of Bim, since activated ERK1/2 phosphorylates Bim and promotes its degradation. Interestingly, TGF-β can induce Bim expression directing its transcription (delayed but sustained expression) or through the expression of the phosphatase MKP2 (rapid response), which deactivates ERK

and prevents its mediated Bim degradation[560]. Therefore, Rac1b-driven downregulation of the TGF-β pathway could also enhance EGFR signalling and favour its resistance to inhibitors (Figure 6, number 7). TCGA analyses showed that RAC1B^{high} tumours expressed lower levels of *MKP2*, once again suggesting the lower activation of the pathway through one of its targeted genes. However, these are aspects that warrant further investigation since non-canonical TGF-β activation was not assessed during my thesis and I did not determine whether *MKP2* was downregulated in my models.

Aside from modulating *SMAD4* expression, an alternative hypothesis might include TGF-β receptor recycling regulation, similar to what I have speculated for EGFR. However, it is shown that clathrin-dependent endocytic recycle of TGF-β would enhance its activity, since endosomal internalisation is required for SMAD activation[576,577]. Thence, it would not explain why Rac1b downregulates the pathway. Alternative internalisation routes exist, such as the endocytosis mediated by lipid-rafts. These facilitate receptor degradation, switching off the signal. TβRs localised at cholesterol-rich lipid rafts associate with the non-canonical/MAPK mediated TGF-β signalling[578]. If an endocytic modulator role was identified for Rac1b, investigating whether it affects TβR internalisation would be of interest. Hints of Rac1b-mediated receptor regulation have also been suggested by Ungefroren *et al*., whose recent data points to the G protein coupled-receptor Par2 as a Rac1b effector for TβRI/ALK5 downregulation[404]. Despite my signalling output being similar to theirs, the cellular function is different. In their cell model, TGF-β induces cell migration and deletion of *RAC1B* translates into a decreased invasion[403,550]. TGF-β response largely depends on the mutational background of the tumoral cell. For instance, *Apc*-only deleted organoids die upon TGF-β treatment (data from the lab). Co-deletion of *TP53* delays death response but virtually no clones survive after 48 hours of TGF-β in the media. However, triple mutant AKP organoids (*Apc*, *K-Ras* and *TP53*) partially overcome the tumour suppressor function of TGF-β and a proportion of cells undergo EMT as a response of TGF-β stimuli (unpublished data). It is yet unknown whether *Rac1b* might inhibit the oncogenic functions of TGF-β or would switch along with it, promoting invasion. Bearing in mind the positive association of *Rac1b* with invasive and metastatic tumours, it might suggest that it would functionally cooperate with TGF-β. The dual response of TGF-β is greatly influenced by the stroma. Most of the TGF-β ligands in advanced tumours are secreted from the stroma and this context-dependent situation might potentially promote its oncogenic switch[184]. The TGF-β-driven oncogenic role of Rac1b might require a stromal component and this might explain why Ungefroren *et al.*

observe a suppressive role of Rac1b during tumour cell invasion in their experiments. Moreover, a large body of evidence involves *Rac1b* in invasive processes, especially through MMP3- and E-cadherin-dependent EMT induction[390,395]. Therefore, the interconnection between Rac1b and TGF-β during invasion awaits further investigations.

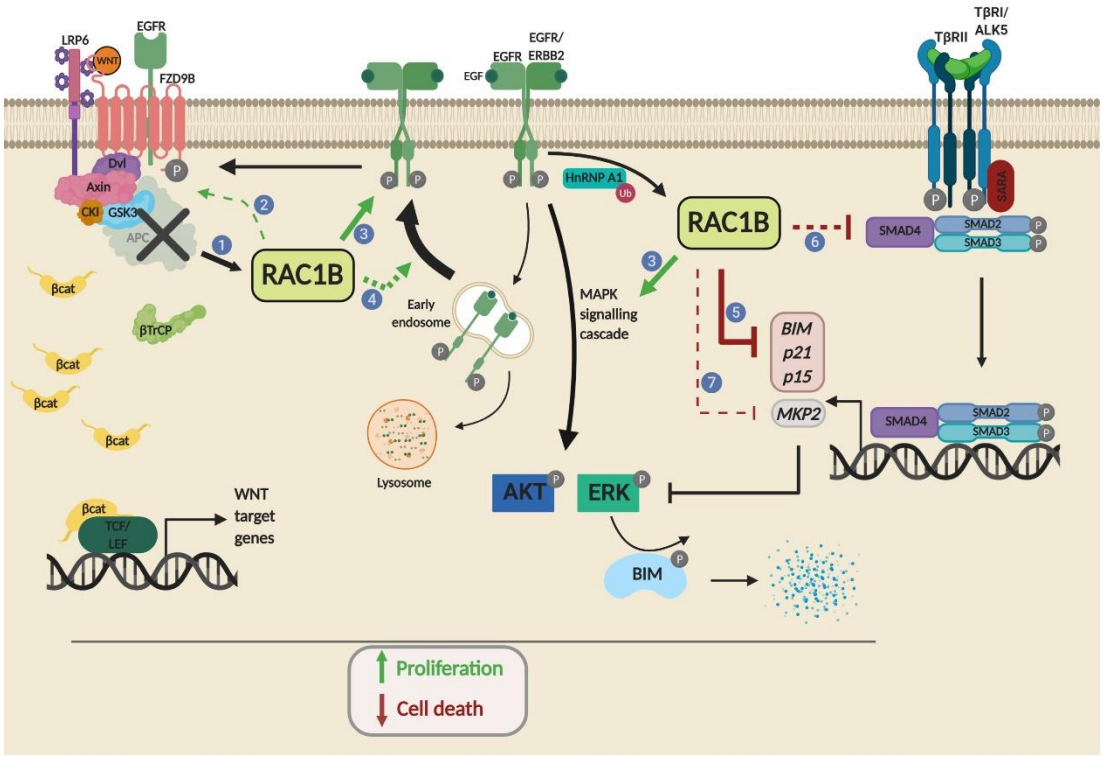


Figure 6: Pro-tumorigenic functions of Rac1b. Upon *Apc* loss, the destruction complex (DC) cannot associate and free cytoplasmic β-catenin translocates to the nucleus, where it binds to TCF/LEF transcription factors and initiates the transcription of WNT target genes. (1) In this scenario, levels of *Rac1b* increase. (2) It has been described that Rac1b might promote WNT signalling through interacting with Dvl. However, I have not observed evidence for this interaction or any other interaction with the DC or WNT receptors. (3) Rac1b is required to maintain EGFR phosphorylation and activation of its downstream pathways, such as AKT and ERK. (4) There are some hints that Rac1b could activate EGFR by promoting its membrane recycling and preventing its lysosomal degradation, although functional evidence is required. Interestingly, EGFR can promote WNT signalling by phosphorylating FZD9B and recruiting the DC to the plasma membrane. EGFR can also potentiate Rac1b expression by targeting for ubiquitylation its negative splicing regulator HnRNP A1. (5) In parallel, Rac1b downregulates TGF-β signalling as observed by the decrease of its target genes *Bim*, *p21* and *p15* and SMAD4. (6) Mechanistically, it could inhibit SMAD4 directly although evidence is required. (7) Moreover, downregulation of Bim could also be achieved by the downregulation of the TGF-β target gene *MKP2*, which would inhibit ERK and prevent Bim degradation. Altogether, through EGFR and WNT activation and TGF-β downregulation, Rac1b expression results in an increased proliferation and a reduced cell death in tumoral cells.
Green and red arrows indicate positive and negative stimulation, respectively. Black arrows correspond to processes described elsewhere. Dashed arrows are processes that require further evidence.

6.4 Therapeutic benefit of targeting Rac1b

Rac1b-deficient mouse organoids and inhibition of *Rac1b* with a specifically designed antisense morpholino in human CRC organoids conferred an enhanced sensitivity to EGFRi in both systems. This is in apparent contradiction with the positive correlation of Rac1b with the CMS2 subtype, which is associated with a good response to EGFR mAbs such as cetuximab. However, I have shown that *RAC1B* expression increases in invasive and metastatic tumours. Further analyses showed that *RAC1B* expression stratifies CMS2 tumours within two subtypes: RAC1B^{low} and good survival or RAC1B^{high} with poor survival and high WNT activity. Interestingly, the later presents a survival as poor as the CMS4, the worst prognostic CMS type. This Rac1b-driven substratification is highly similar to a previous classification attempt conducted by Sadanandam and colleagues[202]. They identified a subtype termed the transit-amplifying group (TA), whose traits largely correspond to those from the CMS2. Interestingly, response to cetuximab stratifies the TA subgroup into cetuximab-sensitive (TA-CS) or cetuximab-resistant (TA-RS) tumours. Given the treatment sensitivity acquired when *Rac1b* is deleted and the survival curve of CMS2-RAC1B^{low}, it seems feasible to think that *RAC1B* expression correlates with tumoral treatment response. At present, analyses are being undertaken to determine whether *RAC1B* expression correlates with expression markers of cetuximab resistance in the TCGA cohort. Altogether, this could indicate that *RAC1B* expression in CMS2 tumours might predict the tumour clinical response to treatment.

EGFR acquired resistance is virtually an unavoidable event. To date, up to 70% of CRC tumours present mutations in members of the EGFR signalling pathway[579]. Even so, resistance is also acquired without the presence of these genetic drivers. Recently, a non-genetic mechanism has been proposed whereby tumours might undergo a transcriptomic switch from CMS2 to CMS4 subtype to develop cetuximab resistance[580]. In line with this, hyperactivation of other signalling pathways might confer resistant traits. For example, it has been shown that concomitantly with EGFR signalling inhibition, WNT signalling increases and its enhanced activity is responsible for acquired cetuximab resistance[171,581]. Both studies suggest that co-treatment with a WNT signalling inhibitor might sensitise resilient tumour cells. Other pathways that become hyperactivated upon cell-autonomous activation of EGFR include Rac1 signalling. Indeed, EGFR promotes Rac1 signalling through Akt-dependent Tiam1 activation, indicating their cross-activity[582]. A number of studies have demonstrated a

synergetic effect when Rac1 activity is targeted along with EGFR[501,502]. However, direct targeting of Rac1 is therapeutically challenging given its ubiquitous location and multimodal function. Alternative strategies might include targeting of GEFs or GAPs to downregulate its activity, but previously I have outlined how their context-dependent function might lead to undesired effects[348,349,532]. Given the specific tumour-related role of Rac1b and the lack of physiological functions, its deletion is foreseen to be well-tolerated with fewer collateral effects. Furthermore, its deletion could simultaneously lead to the decrease of both EGFR and WNT signalling, potentially preventing the resistant response of the WNT pathway. Besides, a positive feedback loop seems to exist between EGFR and Rac1b expression, whereby EGFR promotes exon 3b inclusion[583]. Hence, EGFR targets Rac1b expression to promote tumorigenesis and Rac1b participates in mediating its oncogenic function. An important aspect still to be considered is whether Rac1b might also sensitise tumours mutant for *K-RAS*. *K-RAS* mutation is the main genetic hallmark for EGFR mAbs response and one of the few molecular parameters considered for clinical decision making, since *K-Ras^{mut}* tumours are intrinsically resistant. The pro-apoptotic protein Bim is the main mediator for EGFR-mediated apoptosis and its expression can identify the tumoral response to kinases inhibitors, such as EGFRi or MEKi[584]. Recently, it was shown that the use of a BH3-only mimetics induces cell death to CRC adenoma organoids. Importantly, this mimetic eventually sensitised *K-Ras^{mut}* tumours as well[585]. Given that Rac1b downregulates Bim expression, its deletion might also favour to sensitise *K-Ras^{mut}* tumours. Hence, it is of paramount importance to describe whether this subset of patients (that accounts for up to the half of CRCs) might benefit from *RAC1B* inhibition.

Despite my efforts in determining the potential therapeutic benefit for Rac1b, its inhibition would be unlikely to be beneficial as a single therapy. Although its expression is required for a full oncogenic outcome, its deletion did not completely prevent either tumorigenesis nor EGFR or WNT signal. Moreover, its depletion in models with a higher mutational load such as the APC P53 model did not result in survival improvement. This is consistent with the data obtained with *RAC1B* knock-out in human organoids. While benign organoids presented a growth defect upon Rac1b-morpholino treatment, invasive carcinomas were only affected when co-treated with an EGFRi. This agrees with the current philosophy for anti-cancer treatment, which includes the targeting of more than one gene using more than one inhibitor agent (multi-target drug approach). Therefore, while *Rac1b* inhibition might confer anti-cancer properties, its effect might only be observed clinically

when combined with other agents. I suggest that combined treatment of EGFRi and Rac1b-morpholino might affect tumour performance and potentially avoid or expand the time to relapse. Furthermore, I also provide evidence for a prognostic role of *RAC1B*, whose expression might anticipate therapy response and aid treatment decision.

6.6 Concluding remarks

Since it was discovered 20 years ago, Rac1b has largely been recognised as a tumour-related isoform but *in vivo* investigation of its function was missing. Here, I have provided evidence for a tumour-specific role of Rac1b. Whilst its expression is not required for physiological intestinal homeostasis, Rac1b promotes tumorigenesis. Tumours without Rac1b expression perform poorly, while higher levels of Rac1b enhance cell transformation. Human data indicate a possible role for Rac1b during invasion and metastatic disease, albeit its function awaits *in vivo* evaluation. Importantly, Rac1b might be involved in treatment resistance processes, since depletion of its expression sensitises tumour organoids to EGFR inhibitors. Therefore, these data present Rac1b as a potential druggable target and as an indicator of malignant disease. Further studies are required to validate these findings and feasible translation into the clinic.

Bibliography

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
2. Bowel cancer statistics | Cancer Research UK. Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer. (Accessed: 24th August 2019)
3. Yurgelun, M. B. *et al.* Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. *J. Clin. Oncol.* **35**, 1086–1095 (2017).
4. Brown, K. F. *et al.* The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. *Br. J. Cancer* **118**, 1130–1141 (2018).
5. Bénard, F., Barkun, A. N., Martel, M. & Renteln, D. von. Systematic review of colorectal cancer screening guidelines for average-risk adults: Summarizing the current global recommendations. *World J. Gastroenterol.* **24**, 124–138 (2018).
6. Bouvard, V. *et al.* Carcinogenicity of consumption of red and processed meat. *Lancet Oncol.* **16**, 1599–1600 (2015).
7. Bagnardi, V. *et al.* Alcohol consumption and site-specific cancer risk: a comprehensive dose–response meta-analysis. *Br. J. Cancer* **112**, 580–593 (2015).
8. World Cancer Research Fund / American Institute for Cancer Research. Continuous Update Project Findings and Reports. *September 2017* Available at: https://www.wcrf.org/dietandcancer/cancers. (Accessed: 16th July 2019)
9. Seitz, H. K. & Stickel, F. Molecular mechanisms of alcohol-mediated carcinogenesis. *Nat. Rev. Cancer* **7**, 599–612 (2007).
10. Ma, Y. *et al.* Obesity and Risk of Colorectal Cancer: A Systematic Review of Prospective Studies. *PLoS One* **8**, e53916 (2013).
11. Lauby-Secretan, B. *et al.* Body Fatness and Cancer — Viewpoint of the IARC Working Group. *N. Engl. J. Med.* **375**, 794–798 (2016).
12. Abar, L. *et al.* Height and body fatness and colorectal cancer risk: an update of the WCRF-AICR systematic review of published prospective studies. *Eur. J. Nutr.* **57**, 1701–1720 (2018).
13. Abraham, C. & Cho, J. H. Inflammatory Bowel Disease. *N. Engl. J. Med.* **361**, 2066–2078 (2009).
14. Ananthakrishnan, A. N. Epidemiology and risk factors for IBD. *Nat. Rev. Gastroenterol. Hepatol.* **12**, 205–217 (2015).
15. de Souza, H. S. P. & Fiocchi, C. Immunopathogenesis of IBD: current state of the art. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 13–27 (2016).
16. Lutgens, M. W. M. D. *et al.* Declining Risk of Colorectal Cancer in Inflammatory Bowel Disease. *Inflamm. Bowel Dis.* **19**, 789–799 (2013).
17. Lakatos, P.-L. & Lakatos, L. Risk for colorectal cancer in ulcerative colitis: changes, causes and management strategies. *World J. Gastroenterol.* **14**, 3937–47 (2008).
18. Freeman, H.-J. Colorectal cancer risk in Crohn's disease. *World J. Gastroenterol.* **14**, 1810–1 (2008).
19. de Kort, S. *et al.* Higher risk of colorectal cancer in patients with newly diagnosed diabetes mellitus before the age of colorectal cancer screening initiation. *Sci. Rep.* **7**, 46527 (2017).
20. Limburg, P. J. *et al.* Diabetes mellitus and subsite-specific colorectal cancer risks in the Iowa Women's Health Study. *Cancer Epidemiol. Biomarkers Prev.* **14**, 133–7 (2005).
21. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **0**, 1–9 (2016).
22. *Colorectal cancer Source: Globocan 2018 Number of new cases in 2018, both sexes, all ages*. (2018).
23. Burkitt, D. Related disease—Related cause? *Lancet* **294**, 1229–1231 (1969).
24. Zeng, H., Lazarova, D. L. & Bordonaro, M. Mechanisms linking dietary fiber, gut microbiota

and colon cancer prevention. *World J. Gastrointest. Oncol.* **6**, 41 (2014).
25. Aune, D. *et al.* Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. *BMJ* **343**, d6617 (2011).
26. Chan, A. T. & Giovannucci, E. L. Primary Prevention of Colorectal Cancer. *Gastroenterology* **138**, 2029–2043.e10 (2010).
27. Kong, J. *et al.* Novel role of the vitamin D receptor in maintaining the integrity of the intestinal mucosal barrier. *Am J Physiol Gastrointest Liver Physiol*. **294**, G208-16 (2008).
28. McCullough, M. L. *et al.* Circulating Vitamin D and Colorectal Cancer Risk: An International Pooling Project of 17 Cohorts. *J. Natl. Cancer Inst.* **111**, 158–169 (2019).
29. *Vitamin D and Cancer, a report of the IARC Working Group on Vitamin D, Volume 5 (2008)*.
30. Vaughan-Shaw, P. G. *et al.* Low plasma vitamin D is associated with adverse colorectal cancer survival after surgical resection, independent of systemic inflammatory response. *Gut*. **69**, 103-111 (2019).
31. Yun, J. *et al.* Vitamin C selectively kills KRAS and BRAF mutant colorectal cancer cells by targeting GAPDH. *Sci.* **350**, 1391–6 (2015).
32. Nauman, G., Gray, J. C., Parkinson, R., Levine, M. & Paller, C. J. Systematic Review of Intravenous Ascorbate in Cancer Clinical Trials. *Antioxidants (Basel, Switzerland)* **7**, (2018).
33. Heine-Bröring, R. C. *et al.* Dietary supplement use and colorectal cancer risk: A systematic review and meta-analyses of prospective cohort studies. *Int. J. Cancer* **136**, 2388–2401 (2015).
34. Iamartino, L., Elajnaf, T., Kallay, E. & Schepelmann, M. Calcium-sensing receptor in colorectal inflammation and cancer: Current insights and future perspectives. *World J. Gastroenterol.* **24**, 4119–4131 (2018).
35. Wolin, K. Y., Yan, Y., Colditz, G. A. & Lee, I.-M. Physical activity and colon cancer prevention: a meta-analysis. *Br. J. Cancer* **100**, 611–616 (2009).
36. Boyle, T., Keegel, T., Bull, F., Heyworth, J. & Fritschi, L. Physical Activity and Risks of Proximal and Distal Colon Cancers: A Systematic Review and Meta-analysis. *JNCI J. Natl. Cancer Inst.* **104**, 1548–1561 (2012).
37. Hamoya, T. *et al.* Effects of NSAIDs on the risk factors of colorectal cancer: a mini review. *Genes Environ*. **38**, 6 (2016).
38. Drew, D. A., Cao, Y. & Chan, A. T. Aspirin and colorectal cancer: the promise of precision chemoprevention. *Nat. Rev. Cancer* **16**, 173–186 (2016).
39. Din, F. V. N. *et al.* Aspirin Inhibits mTOR Signaling, Activates AMP-Activated Protein Kinase, and Induces Autophagy in Colorectal Cancer Cells. *Gastroenterology* **142**, 1504–1515.e3 (2012).
40. Rothwell, P. M. *et al.* Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. *Lancet* **376**, 1741–1750 (2010).
41. Rothwell, P. M. *et al.* Effects of aspirin on risks of vascular events and cancer according to bodyweight and dose: analysis of individual patient data from randomised trials. *Lancet* **392**, 387–399 (2018).
42. Bibbins-Domingo, K. Aspirin Use for the Primary Prevention of Cardiovascular Disease and Colorectal Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann. Intern. Med.* **164**, 836 (2016).
43. Vasen, H. F. A., Tomlinson, I. & Castells, A. Clinical management of hereditary colorectal cancer syndromes. *Nat. Rev. Gastroenterol. Hepatol.* **12**, 88–97 (2015).
44. Bansidhar, B. J. Extracolonic manifestations of lynch syndrome. *Clin. Colon Rectal Surg.* **25**, 103–10 (2012).
45. Lynch, H. T. & de la Chapelle, A. Hereditary Colorectal Cancer. *N. Engl. J. Med.* **348**, 919–932 (2003).
46. Bonadona, V. *et al.* Cancer Risks Associated With Germline Mutations in MLH1, MSH2, and MSH6 genes in Lynch Syndrome. *JAMA* **305**, 2304 (2011).
47. Vasen, H. F. A. *et al.* Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. *Gut* **62**, 812–823 (2013).
48. de la Chapelle, A. Genetic predisposition to colorectal cancer. *Nat. Rev. Cancer* **4**, 769–780

(2004).
49. Lorans, M., Dow, E., Macrae, F. A., Winship, I. M. & Buchanan, D. D. Update on Hereditary Colorectal Cancer: Improving the Clinical Utility of Multigene Panel Testing. *Clin. Colorectal Cancer* **17**, e293–e305 (2018).
50. Leoz, M. L., Carballal, S., Moreira, L., Ocaña, T. & Balaguer, F. The genetic basis of familial adenomatous polyposis and its implications for clinical practice and risk management. *Appl. Clin. Genet.* **8**, 95–107 (2015).
51. Knudsen, A. L., Bisgaard, M. L. & Bülow, S. *Attenuated familial adenomatous polyposis (AFAP). A review of the literature*. *Familial Cancer* **2**, (2003).
52. Ricciardiello, L., Ahnen, D. J. & Lynch, P. M. Chemoprevention of hereditary colon cancers: time for new strategies. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 352–361 (2016).
53. Al-Tassan, N. *et al.* Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nat. Genet.* **30**, 227–232 (2002).
54. Sieber, O. M. *et al.* Multiple Colorectal Adenomas, Classic Adenomatous Polyposis, and Germ-Line Mutations in MYH. *N. Engl. J. Med.* **348**, 791–799 (2003).
55. Nielsen, M. *et al.* Germline mutations in APC and MUTYH are responsible for the majority of families with attenuated familial adenomatous polyposis. *Clin. Genet.* **71**, 427–433 (2007).
56. Farrington, S. M. *et al.* Germline Susceptibility to Colorectal Cancer Due to Base-Excision Repair Gene Defects. *Am. J. Hum. Genet.* **77**, 112–119 (2005).
57. Jelsig, A. M. *et al.* Hamartomatous polyposis syndromes: a review. *Orphanet J. Rare Dis.* **9**, 101 (2014).
58. Islam, R. S., Patel, N. C., Lam-Himlin, D. & Nguyen, C. C. Gastric polyps: a review of clinical, endoscopic, and histopathologic features and management decisions. *Gastroenterol. Hepatol. (N. Y).* **9**, 640–51 (2013).
59. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
60. Rowan, A. J. *et al.* APC mutations in sporadic colorectal tumors: A mutational "hotspot" and interdependence of the "two hits". *Proc. Natl. Acad. Sci.* **97**, 3352–3357 (2000).
61. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. Regulation of intracellular beta-catenin levels by the adenomatous polyposis coli (APC) tumor-suppressor protein. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 3046–50 (1995).
62. Huelsken, J., Behrens, J., Jeon, S. H., Lee, J.-S. & Choi, K.-Y. The Wnt signalling pathway. *J. Cell Sci.* **115**, 3977–8 (2002).
63. Nakamura, M., Zhou, X. Z. & Lu, K. P. Critical role for the EB1 and APC interaction in the regulation of microtubule polymerization. *Curr. Biol.* **11**, 1062–7 (2001).
64. Green, R. A., Wollman, R. & Kaplan, K. B. APC and EB1 function together in mitosis to regulate spindle dynamics and chromosome alignment. *Mol. Biol. Cell* **16**, 4609–22 (2005).
65. Green, R. A. & Kaplan, K. B. Chromosome instability in colorectal tumor cells is associated with defects in microtubule plus-end attachments caused by a dominant mutation in APC. *J. Cell Biol.* **163**, 949–61 (2003).
66. Hadjihannas, M. V. *et al.* Aberrant Wnt/beta-catenin signaling can induce chromosomal instability in colon cancer. *Proc. Natl. Acad. Sci.* **103**, 10747–10752 (2006).
67. Rodriguez, J. *et al.* Chromosomal Instability Correlates with Genome-wide DNA Demethylation in Human Primary Colorectal Cancers. *Cancer Res.* **66**, 8462–9468 (2006).
68. Weisenberger, D. J. *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat. Genet.* **38**, 787–793 (2006).
69. Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. RAS oncogenes: weaving a tumorigenic web. *Nat. Rev. Cancer* **11**, 761–774 (2011).
70. Eric R. Fearon. Molecular Genetics of Colorectal Cancer. *Annu Rev Pathol*. **6**, 479-507 (2011)
71. Rajagopalan, H. *et al.* RAF/RAS oncogenes and mismatch-repair status. *Nature* **418**, 934–934 (2002).
72. Feng, Y. *et al.* Mutant Kras Promotes Hyperplasia and Alters Differentiation in the Colon Epithelium but Does Not Expand the Presumptive Stem Cell Pool. *YGAST* **141**, 1003–1013.e10

(2011).
73. Yun, J. *et al.* Glucose Deprivation Contributes to the Development of KRAS Pathway Mutations in Tumor Cells. *Science (80-.).* **325**, 1555–1559 (2009).
74. Hutton, J. E. *et al.* Oncogenic KRAS and BRAF Drive Metabolic Reprogramming in Colorectal Cancer. *Mol. Cell. Proteomics* **15**, 2924–2938 (2016).
75. Leslie, A., Carey, F. A., Pratt, N. R. & Steele, R. J. C. The colorectal adenoma-carcinoma sequence. *Br. J. Surg.* **89**, 845–860 (2002).
76. Mehlen, P. & Fearon, E. R. Role of the dependence receptor DCC in colorectal cancer pathogenesis. *J. Clin. Oncol.* **22**, 3420–8 (2004).
77. Furne, C., Rama, N., Corset, V., Chédotal, A. & Mehlen, P. Netrin-1 is a survival factor during commissural neuron navigation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14465–70 (2008).
78. Fazeli, A. *et al.* Phenotype of mice lacking functional Deleted in colorectal cancer (Dec) gene. *Nature* **386**, 796–804 (1997).
79. Castets, M. *et al.* DCC constrains tumour progression via its dependence receptor activity. *Nature* **482**, 534–537 (2012).
80. Takaku, K. *et al.* Gastric and duodenal polyps in Smad4 (Dpc4) knockout mice. *Cancer Res.* **59**, 6113–7 (1999).
81. Shi, Y. & Massagué, J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685–700 (2003).
82. Levine, A. J. & Oren, M. The first 30 years of p53: growing ever more complex. *Nat. Rev. Cancer* **9**, 749–758 (2009).
83. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
84. López, I. *et al.* Different mutation profiles associated to P53 accumulation in colorectal cancer. *Gene* **499**, 81–87 (2012).
85. Jones, S. *et al. Comparative lesion sequencing provides insights into tumor evolution*. (2008).
86. Carracedo, A. & Pandolfi, P. P. The PTEN–PI3K pathway: of feedbacks and cross-talks. *Oncogene* **27**, 5527–5541 (2008).
87. Chalhoub, N. & Baker, S. J. PTEN and the PI3-Kinase Pathway in Cancer PI3K: phosphatidylinositol 3-kinase. *Annu Rev Pathol*. 4, 127-150 (2008).
88. Fearon, E. R. Molecular Genetics of Colorectal Cancer. *Annu. Rev. Pathol. Mech. Dis.* **6**, 479–507 (2011).
89. Marsh, V. *et al.* Epithelial Pten is dispensable for intestinal homeostasis but suppresses adenoma development and progression after Apc mutation. *Nature Genetics*. **40**, 1436-1444 (2008)
90. Samuels, Y. *et al.* High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**, 554 (2004).
91. Longacre, T. A. & Fenoglio-Preiser, C. M. Mixed hyperplastic adenomatous polyps/serrated adenomas. A distinct form of colorectal neoplasia. *Am. J. Surg. Pathol.* **14**, 524–37 (1990).
92. Szylberg, A., Janiczek, M., Popiel, A. & Marszabek, A. Serrated Polyps and Their Alternative Pathway to the Colorectal Cancer: A Systematic Review. *Gastroenterol Res Pract.* 2015-573814 (2015)
93. De Palma, F. *et al.* The Molecular Hallmarks of the Serrated Pathway in Colorectal Cancer. *Cancers (Basel).* **11**, 1017 (2019).
94. Rex, D. K. *et al.* Serrated lesions of the colorectum: review and recommendations from an expert panel. *Am. J. Gastroenterol.* **107**, 1315–29 (2012).
95. Jass, J. R. Serrated adenoma of the colorectum and the DNA-methylator phenotype. *Nat. Clin. Pract. Oncol.* **2**, 398–405 (2005).
96. Toyota, M. *et al.* CpG island methylator phenotype in colorectal cancer. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 8681–6 (1999).
97. Park, S.-J. *et al.* Frequent CpG island methylation in serrated adenomas of the colorectum. *Am. J. Pathol.* **162**, 815–22 (2003).
98. Bettington, M. *et al.* The serrated pathway to colorectal carcinoma: current concepts and challenges. *Histopathology* **62**, 367–386 (2013).

99. Fennell, L. *et al.* Integrative Genome-Scale DNA Methylation Analysis of a Large and Unselected Cohort Reveals 5 Distinct Subtypes of Colorectal Adenocarcinomas. *Cell. Mol. Gastroenterol. Hepatol.* **8**, 269–290 (2019).
100. Spring, K. J. *et al.* High Prevalence of Sessile Serrated Adenomas With BRAF Mutations: A Prospective Study of Patients Undergoing Colonoscopy. *Gastroenterology.* **131**, 1400-1407 (2006).
101. Borowsky, J. *et al.* The role of APC in WNT pathway activation in serrated neoplasia. *Mod. Pathol.* **31**, 495–504 (2018).
102. Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–216 (2015).
103. Khambata-Ford, S. *et al.* Expression of epiregulin and amphiregulin and K-ras mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab. *J. Clin. Oncol.* **25**, 3230–7 (2007).
104. Misale, S. *et al.* Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532–536 (2012).
105. Diaz Jr, L. A. *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537–540 (2012).
106. Hu, Z. *et al.* Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat. Genet.* **51**, 1113–1122 (2019).
107. Williams, M. J., Werner, B., Barnes, C. P., Graham, T. A. & Sottoriva, A. Identification of neutral tumor evolution across cancer types. *Nat. Genet.* **48**, 238–244 (2016).
108. Ling, S. *et al.* Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6496-505 (2015).
109. Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
110. Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science.* **355**, eaaf8399 (2017).
111. Tarabichi, M. *et al.* Neutral tumor evolution? *Nat. Genet.* **50**, 1630–1633 (2018).
112. Williams, M. J. *et al.* Reply to 'Revisiting signatures of neutral tumor evolution in the light of complexity of cancer genomic data'. *Nat. Genet.* **50**, 1628–1630 (2018).
113. Almendro, V. *et al.* Inference of Tumor Evolution during Chemotherapy by Computational Modeling and In Situ Analysis of Genetic and Phenotypic Cellular Diversity. *Cell Rep.* **6**, 514–527 (2014).
114. Jackstadt, R. & Sansom, O. J. Mouse models of intestinal cancer. *J. Pathol.* **238**, 141–151 (2016).
115. Drost, J. *et al.* Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43–47 (2015).
116. Matano, M. *et al.* Modeling colorectal cancer using CRISPR-Cas9–mediated engineering of human intestinal organoids. *Nat. Med.* **21**, 256–262 (2015).
117. Fumagalli, A. *et al.* Genetic dissection of colorectal cancer progression by orthotopic transplantation of engineered cancer organoids. *Proc Natl Acad Sci USA*. **114**, E2357-E2364 (2017)
118. Helander, H. F. & Fändriks, L. Surface area of the digestive tract – revisited. *Scand. J. Gastroenterol.* **49**, 681–689 (2014).
119. Collins, J. T. & Badireddy, M. *Anatomy, Abdomen and Pelvis, Small Intestine*. *StatPearls* (2019).
120. Gehart, H. & Clevers, H. Tales from the crypt: new insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 19–34 (2019).
121. Barker, N. Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. *Nat. Rev. Mol. Cell Biol.* **15**, 19–33 (2014).
122. Ireland, H., Houghton, C., Howard, L. & Winton, D. J. Cellular inheritance of a Cre-activated reporter gene to determine paneth cell longevity in the murine small intestine. *Dev. Dyn.* **233**, 1332–1336 (2005).

123. Crawley, S. W., Mooseker, M. S. & Tyska, M. J. Shaping the intestinal brush border. *J. Cell Biol.* **207**, 441–451 (2014).
124. Ohno, H. Intestinal M cells. *J. Biochem.* **159**, 151–160 (2016).
125. Bjerknes, M. *et al.* Origin of the brush cell lineage in the mouse intestinal epithelium. *Dev. Biol.* **362**, 194–218 (2012).
126. Sternini, C., Anselmi, L. & Rozengurt, E. Enteroendocrine cells: a site of 'taste' in gastrointestinal chemosensing. *Curr. Opin. Endocrinol. Diabetes Obes.* **15**, 73–78 (2008).
127. van der Flier, L. G. & Clevers, H. Stem Cells, Self-Renewal, and Differentiation in the Intestinal Epithelium. *Annu. Rev. Physiol.* **71**, 241–260 (2009).
128. Sasaki, N. *et al.* Reg4+ deep crypt secretory cells function as epithelial niche for Lgr5+ stem cells in colon. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5399-407 (2016).
129. Potten, C. S., Kovacs, L. & Hamilton, E. Continuous labelling studies on mouse skin and intestine. *Cell Tissue Kinet.* **7**, 271–83 (1974).
130. Cheng, H. & Leblond, C. P. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine I. Columnar cell. *Am. J. Anat.* **141**, 461–479 (1974).
131. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007).
132. Tetteh, P. W., Farin, H. F. & Clevers, H. Plasticity within stem cell hierarchies in mammalian epithelia. *Trends Cell Biol.* **25**, 100–108 (2015).
133. Buczacki, S. J. A. *et al.* Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature* **495**, 65–69 (2013).
134. Tetteh, P. W. *et al.* Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. *Cell Stem Cell* **18**, 203–213 (2016).
135. van Es, J. H. *et al.* Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. *Nat. Cell Biol.* **14**, 1099–1104 (2012).
136. Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
137. Rodríguez-Colman, M. J. *et al.* Interplay between metabolic identities in the intestinal crypt supports stem cell function. *Nature* **543**, 424–427 (2017).
138. Kim, T.-H. *et al.* Broadly permissive intestinal chromatin underlies lateral inhibition and cell plasticity. *Nature* **506**, 511–515 (2014).
139. Tata, P. R. *et al.* Dedifferentiation of committed epithelial cells into stem cells in vivo. *Nature* **503**, 218–223 (2013).
140. Tarlow, B. D. *et al.* Bipotential Adult Liver Progenitors Are Derived from Chronically Injured Mature Hepatocytes. *Cell Stem Cell* **15**, 605–618 (2014).
141. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
142. Pinto, D., Gregorieff, A., Begthel, H. & Clevers, H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes Dev.* **17**, 1709–1713 (2003).
143. Kuhnert, F. *et al.* Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. *Proc Natl Acad Sci USA.* **101**, 266-71 (2004).
144. Wilkins, J. A. & Sansom, O. J. C-Myc Is a Critical Mediator of the Phenotypes of Apc Loss in the Intestine. *Cancer Res.* **68**, 4963–4966 (2008).
145. Valenta, T. *et al.* Wnt Ligands Secreted by Subepithelial Mesenchymal Cells Are Essential for the Survival of Intestinal Stem Cells and Gut Homeostasis. *Cell Rep.* **15**, 911–918 (2016).
146. Greicius, G. *et al.* PDGFRα+ pericryptal stromal cells are the critical source of Wnts and RSPO3 for murine intestinal stem cells in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E3173–E3181 (2018).
147. Farin, H. F. *et al.* Visualization of a short-range Wnt gradient in the intestinal stem-cell niche. *Nature* **530**, 340–343 (2016).
148. Carmon, K. S., Gong, X., Lin, Q., Thomas, A. & Liu, Q. R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/ -catenin signaling. *Proc. Natl. Acad. Sci.* **108**, 11452–11457 (2011).

149. Batlle, E. *et al.* Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* **111**, 251–63 (2002).
150. Wong, V. W. Y. *et al.* Lrig1 controls intestinal stem-cell homeostasis by negative regulation of ErbB signalling. *Nat. Cell Biol.* **14**, 401–408 (2012).
151. Basak, O. *et al.* Induced Quiescence of Lgr5+ Stem Cells in Intestinal Organoids Enables Differentiation of Hormone-Producing Enteroendocrine Cells. *Cell Stem Cell* **20**, 177–190.e4 (2017).
152. Haramis, A.-P. G. *et al.* De Novo Crypt Formation and Juvenile Polyposis on BMP Inhibition in Mouse Intestine. *Science (80-.).* **303**, 1684–1686 (2004).
153. Kosinski, C. *et al.* Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. *Proc. Natl. Acad. Sci.* **104**, 15418–15423 (2007).
154. Sancho, R., Cremona, C. A. & Behrens, A. Stem cell and progenitor fate in the mammalian intestine: Notch and lateral inhibition in homeostasis and disease. *EMBO Rep.* **16**, 571–581 (2015).
155. Sanchez-Vega, F. *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321–337.e10 (2018).
156. Najdi, R. *et al.* A uniform human Wnt expression library reveals a shared secretory pathway and unique signaling activities. *Differentiation* **84**, 203–213 (2012).
157. Alok, A. *et al.* Wnt proteins synergize to activate β-catenin signaling. *J. Cell Sci.* **130**, 1532–1544 (2017).
158. Nam, J.-S., Turcotte, T. J., Smith, P. F., Choi, S. & Yoon, J. K. Mouse cristin/R-spondin family proteins are novel ligands for the Frizzled 8 and LRP6 receptors and activate beta-catenin-dependent gene expression. *J. Biol. Chem.* **281**, 13247–57 (2006).
159. Yan, K. S. *et al.* Non-equivalence of Wnt and R-spondin ligands during Lgr5+ intestinal stem-cell self-renewal. *Nature* **545**, 238–242 (2017).
160. Fiedler, M., Mendoza-Topaz, C., Rutherford, T. J., Mieszczanek, J. & Bienz, M. Dishevelled interacts with the DIX domain polymerization interface of Axin to interfere with its function in down-regulating β-catenin. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1937–42 (2011).
161. Lau, W. de *et al.* Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. *Nat. 2011 4767360* **476**, 293–297 (2011).
162. Hao, H.-X. *et al.* ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature* **485**, 195–200 (2012).
163. Jiang, X., Charlat, O., Zamponi, R., Yang, Y. & Cong, F. Dishevelled promotes Wnt receptor degradation through recruitment of ZNRF3/RNF43 E3 ubiquitin ligases. *Mol. Cell* **58**, 522–33 (2015).
164. Li, V. S. W. *et al.* Wnt Signaling through Inhibition of β-Catenin Degradation in an Intact Axin1 Complex. *Cell* **149**, 1245–1256 (2012).
165. Macdonald, B. T., Tamai, K. & He, X. Developmental Cell Review Wnt/b-Catenin Signaling: Components, Mechanisms, and Diseases. *Dev. Cell* **17**, 9–26
166. Van der Flier, L. G. *et al.* The Intestinal Wnt/TCF Signature. *Gastroenterology* **132**, 628–632 (2007).
167. Sanz-Pamplona, R. *et al.* Exome Sequencing Reveals AMER1 as a Frequently Mutated Gene in Colorectal Cancer. *Clin. Cancer Res.* **21**, 4709–4718 (2015).
168. Bafico, A., Liu, G., Goldin, L., Harris, V. & Aaronson, S. A. An autocrine mechanism for constitutive Wnt pathway activation in human cancer cells. *Cancer Cell* **6**, 497–506 (2004).
169. Huels, D. J. *et al.* Wnt ligands influence tumour initiation by controlling the number of intestinal stem cells. *Nat. Commun.* **9**, 1132 (2018).
170. Suzuki, H. *et al.* Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. *Nat. Genet.* **36**, 417–422 (2004).
171. Zhan, T. *et al.* MEK inhibitors activate Wnt signalling and induce stem cell plasticity in colorectal cancer. *Nat. Commun.* **10**, 2197 (2019).
172. Zhang, W. & Liu, H. T. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res.* **12**, 9–18 (2002).

173. Meloche, S. & Pouysségur, J. The ERK1/2 mitogen-activated protein kinase pathway as a master regulator of the G1- to S-phase transition. *Oncogene* **26**, 3227–3239 (2007).
174. Seki, E., Brenner, D. A. & Karin, M. A Liver Full of JNK: Signaling in Regulation of Cell Function and Disease Pathogenesis, and Clinical Approaches. *Gastroenterology* **143**, 307–320 (2012).
175. Zarubin, T. & Han, J. Activation and signaling of the p38 MAP kinase pathway. *Cell Res.* **15**, 11–18 (2005).
176. Osaki, L. H. & Gama, P. MAPKs and Signal Transduction in the Control of Gastrointestinal Epithelial Cell Proliferation and Differentiation. *Int. J. Mol. Sci* **14**, 10143–10161 (2013).
177. Aliaga, J. C., Deschênes, C., Beaulieu, J.-F., Calvo, E. L. & Rivard, N. Requirement of the MAP kinase cascade for cell cycle progression and differentiation of human intestinal cells. *Am. J. Physiol. Liver Physiol.* **277**, G631–G641 (1999).
178. Infante, J. R. *et al.* Safety, pharmacokinetic, pharmacodynamic, and efficacy data for the oral MEK inhibitor trametinib: a phase 1 dose-escalation trial. *Lancet. Oncol.* **13**, 773–81 (2012).
179. Massagué, J. TGFβ signalling in context. *Nat. Rev. Mol. Cell Biol.* **13**, 616–630 (2012).
180. Zhang, Y. E. Non-Smad Signaling Pathways of the TGF-β Family. *Cold Spring Harb. Perspect. Biol.* **9**, a022129 (2017).
181. Heldin, C.-H., Ne Landströ, M. & Moustakas, A. Mechanism of TGF-b signaling to growth arrest, apoptosis, and epithelial-mesenchymal transition. *Curr. Opin. Cell Biol.* **21**, 166–176 (2009).
182. Inman, G. J. Switching TGFβ from a tumor suppressor to a tumor promoter. *Curr. Opin. Genet. Dev.* **21**, 93–99 (2011).
183. cBioPortal for Cancer Genomics. Available at: https://www.cbioportal.org/. (Accessed: 27th May 2019)
184. Calon, A. *et al.* Dependency of Colorectal Cancer on a TGF-β-Driven Program in Stromal Cells for Metastasis Initiation. *Cancer Cell* **22**, 571–584 (2012).
185. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
186. Green, D. R. & Kroemer, G. Cytoplasmic functions of the tumour suppressor p53. *Nature* **458**, 1127–1130 (2009).
187. Menendez, D., Inga, A. & Resnick, M. A. The expanding universe of p53 targets. *Nat. Rev. Cancer* **9**, 724–737 (2009).
188. Linares, L. K., Hengstermann, A., Ciechanover, A., Muller, S. & Scheffner, M. HdmX stimulates Hdm2-mediated ubiquitination and degradation of p53. *Proc. Natl. Acad. Sci.* **100**, 12009–12014 (2003).
189. Karen H. Vousden and Carol Prives. Blinded by the Light: The Growing Complexity of p53.
190. Lane, D. P. p53, guardian of the genome. *Nature* **358**, 15–16 (1992).
191. Hussain, S. P. *et al.* Increased p53 mutation load in noncancerous colon tissue from ulcerative colitis: a cancer-prone chronic inflammatory disease. *Cancer Res.* **60**, 3333–7 (2000).
192. BLAGOSKLONNY, M. V. p53 from complexity to simplicity: mutant p53 stabilization, gain-of-function, and dominant-negative effect. *FASEB J.* **14**, 1901–1907 (2000).
193. Muller, P. A. J. *et al.* Mutant p53 drives invasion by promoting integrin recycling. *Cell* **139**, 1327–41 (2009).
194. Muller, P. A. J. & Vousden, K. H. Mutant p53 in Cancer: New Functions and Therapeutic Opportunities. *Cancer Cell* **25**, 304–317 (2014).
195. Nakayama, M. *et al.* Intestinal cancer progression by mutant p53 through the acquisition of invasiveness associated with complex glandular formation. *Oncogene* **36**, 5885–5896 (2017).
196. Boettcher, S. *et al.* A dominant-negative effect drives selection of TP53 missense mutations in myeloid malignancies. *Science* **365**, 599–604 (2019).
197. Robertson, A. G. *et al.* Integrative Analysis Identifies Four Molecular and Clinical Subsets in Uveal Melanoma. *Cancer Cell* **32**, 204–220.e15 (2017).
198. Hutter, C. & Zenklusen, J. C. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell* **173**, 283–285 (2018).
199. Boland, C. R. & Goel, A. Microsatellite Instability in Colorectal Cancer. *Gastroenterology* **138**,

2073–2087.e3 (2010).
200. Donehower, L. A. *et al.* MLH1 -silenced and non-silenced subgroups of hypermutated colorectal carcinomas have distinct mutational landscapes. *J. Pathol.* **229**, 99–110 (2013).
201. Network, T. C. G. A. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
202. Sadanandam, A. *et al.* A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat. Med.* **19**, 619–625 (2013).
203. Marisa, L. *et al.* Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* **10**, e1001453 (2013).
204. De Sousa E Melo, F. *et al.* Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat. Med.* **19**, 614–618 (2013).
205. Isella, C. *et al.* Stromal contribution to the colorectal cancer transcriptome. *Nat. Genet.* **47**, 312–319 (2015).
206. Calon, A. *et al.* Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.* **47**, 320–329 (2015).
207. Benson, A. B. *et al.* NCCN Guidelines Insights: Colon Cancer, Version 2.2018. *J. Natl. Compr. Cancer Netw.* **16**, 359–369 (2018).
208. Fontana, E., Eason, K., Cervantes, A., Salazar, R. & Sadanandam, A. Context matters—consensus molecular subtypes of colorectal cancer as biomarkers for clinical trials. *Ann. Oncol.* **30**, 520–527 (2019).
209. Dunne, P. D. *et al.* Challenging the Cancer Molecular Stratification Dogma: Intratumoral Heterogeneity Undermines Consensus Molecular Subtypes and Potential Diagnostic Value in Colorectal Cancer. *Clin. Cancer Res.* **22**, 4095–4104 (2016).
210. Punt, C. J. A., Koopman, M. & Vermeulen, L. From tumour heterogeneity to advances in precision treatment of colorectal cancer. *Nat. Rev. Clin. Oncol.* **14**, 235–246 (2017).
211. Loupakis, F. *et al.* Initial Therapy with FOLFOXIRI and Bevacizumab for Metastatic Colorectal Cancer. *N. Engl. J. Med.* **371**, 1609–1618 (2014).
212. Hoff, P. M. *et al.* Comparison of Oral Capecitabine Versus Intravenous Fluorouracil Plus Leucovorin as First-Line Treatment in 605 Patients With Metastatic Colorectal Cancer: Results of a Randomized Phase III Study. *J. Clin. Oncol.* **19**, 2282–2292 (2001).
213. Falcone, A. *et al.* Phase III trial of infusional fluorouracil, leucovorin, oxaliplatin, and irinotecan (FOLFOXIRI) compared with infusional fluorouracil, leucovorin, and irinotecan (FOLFIRI) as first-line treatment for metastatic colorectal cancer: the Gruppo Oncologico Nord Ovest. *J. Clin. Oncol.* **25**, 1670–6 (2007).
214. André, T. *et al.* Adjuvant Fluorouracil, Leucovorin, and Oxaliplatin in Stage II to III Colon Cancer: Updated 10-Year Survival and Outcomes According to *BRAF* Mutation and Mismatch Repair Status of the MOSAIC Study. *J. Clin. Oncol.* **33**, 4176–4187 (2015).
215. Karapetis, C. S. *et al.* K-ras Mutations and Benefit from Cetuximab in Advanced Colorectal Cancer. *N. Engl. J. Med.* **359**, 1757–1765 (2008).
216. Corcoran, R. B. *et al.* EGFR-Mediated Reactivation of MAPK Signaling Contributes to Insensitivity of BRAF Mutant Colorectal Cancers to RAF Inhibition with Vemurafenib. *Cancer Discov.* **2**, 227–235 (2012).
217. Sargent, D. J. *et al.* Defective Mismatch Repair As a Predictive Marker for Lack of Efficacy of Fluorouracil-Based Adjuvant Therapy in Colon Cancer. *J. Clin. Oncol.* **28**, 3219–3226 (2010).
218. Douillard, J.-Y. *et al.* Randomized, Phase III Trial of Panitumumab With Infusional Fluorouracil, Leucovorin, and Oxaliplatin (FOLFOX4) Versus FOLFOX4 Alone As First-Line Treatment in Patients With Previously Untreated Metastatic Colorectal Cancer: The PRIME Study. *J. Clin. Oncol.* **28**, 4697–4705 (2010).
219. Van Cutsem, E. *et al.* Cetuximab and Chemotherapy as Initial Treatment for Metastatic Colorectal Cancer. *N. Engl. J. Med.* **360**, 1408–1417 (2009).
220. Hurwitz, H. *et al.* Bevacizumab plus Irinotecan, Fluorouracil, and Leucovorin for Metastatic Colorectal Cancer. *N. Engl. J. Med.* **350**, 2335–2342 (2004).
221. Van Cutsem, E., Cervantes, A., Nordlinger, B., Arnold, D. & ESMO Guidelines Working Group. Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and

follow-up. *Ann. Oncol.* **25**, iii1-iii9 (2014).
222. Rojas, A. M., Fuentes, G., Rausell, A. & Valencia, A. The Ras protein superfamily: evolutionary tree and role of conserved amino acids. *J. Cell Biol.* **196**, 189–201 (2012).
223. Colicelli, J. Human RAS superfamily proteins and related GTPases. *Sci. STKE* **2004**, RE13 (2004).
224. Herrmann, C. Ras-effector interactions: after one decade. *Curr. Opin. Struct. Biol.* **13**, 122–9 (2003).
225. Hodge, R. G. & Ridley, A. J. Regulating Rho GTPases and their regulators. *Nat. Rev. Mol. Cell Biol.* **17**, 496–510 (2016).
226. Eisenberg, S. *et al.* The role of palmitoylation in regulating Ras localization and function. *Biochem. Soc. Trans.* **41**, 79–83 (2013).
227. Wennerberg, K., Rossman, K. L. & Der, C. J. The Ras superfamily at a glance. *J. Cell Sci.* **118**, 843–846 (2005).
228. Nethe, M. & Hordijk, P. L. The role of ubiquitylation and degradation in RhoGTPase signalling. *J. Cell Sci.* **123**, 4011–4018 (2010).
229. Cox, A. D. & Der, C. J. Ras history. *Small GTPases* **1**, 2–27 (2010).
230. Grosshans, B. L., Ortiz, D. & Novick, P. Rabs and their effectors: Achieving specificity in membrane traffic. *Proc. Natl. Acad. Sci.* **103**, 11821–11827 (2006).
231. Donaldson, J. G. & Jackson, C. L. ARF family G proteins and their regulators: roles in membrane transport, development and disease. *Nat. Rev. Mol. Cell Biol.* **12**, 362–375 (2011).
232. Clarke, P. R. & Zhang, C. Ran GTPase: a master regulator of nuclear structure and function during the eukaryotic cell division cycle? *Trends Cell Biol.* **11**, 366–71 (2001).
233. Heasman, S. J. & Ridley, A. J. Mammalian Rho GTPases: new insights into their functions from in vivo studies. *Nat. Rev. Mol. Cell Biol.* **9**, 690–701 (2008).
234. Vega, F. M. & Ridley, A. J. Rho GTPases in cancer cell biology. *FEBS Lett.* **582**, 2093–2101 (2008).
235. Madaule, P. & Axel, R. A novel ras-related gene family. *Cell* **41**, 31–40 (1985).
236. Didsbury, J., Weber, R. F., Bokoch, G. M., Evans, T. & Snyderman, R. rac, a novel ras-related family of proteins that are botulinum toxin substrates. *J. Biol. Chem.* **264**, 16378–82 (1989).
237. Munemitsu, S. *et al.* Molecular cloning and expression of a G25K cDNA, the human homolog of the yeast cell cycle gene CDC42. *Mol. Cell. Biol.* **10**, 5977–82 (1990).
238. Schmidt, A. Guanine nucleotide exchange factors for Rho GTPases: turning on the switch. *Genes Dev.* **16**, 1587–1609 (2002).
239. Cherfils, J. & Zeghouf, M. Regulation of Small GTPases by GEFs, GAPs, and GDIs. *Physiol. Rev.* **93**, 269–309 (2013).
240. Goitre, L., Trapani, E., Trabalzini, L. & Retta, S. F. The Ras Superfamily of Small GTPases: The Unlocked Secrets. in *Methods in molecular biology* **1120**, 1–18 (2014).
241. Ridley, A. J. & Hall, A. The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell* **70**, 389–399 (1992).
242. Wheeler, A. & Ridley, A. Why three Rho proteins? RhoA, RhoB, RhoC, and cell motility. *Exp. Cell Res.* **301**, 43–49 (2004).
243. Ridley, A. J. *et al.* Cell Migration: Integrating Signals from Front to Back. *Science.* **302**, 1704–1709 (2003).
244. RIDLEY, A. J. RhoA, RhoB and RhoC have different roles in cancer cell migration. *J. Microsc.* **251**, 242–249 (2013).
245. Schaefer, A., Reinhard, N. R. & Hordijk, P. L. Toward understanding RhoGTPase specificity: structure, function and local activation. *Small GTPases* **5**, e968004 (2014).
246. Clark, E. A., Golub, T. R., Lander, E. S. & Hynes, R. O. Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* **406**, 532–535 (2000).
247. Wherlock, M., Gampel, A., Futter, C. & Mellor, H. Farnesyltransferase inhibitors disrupt EGF receptor traffic through modulation of the RhoB GTPase. *J. Cell Sci.* **117**, 3221–3231 (2004).
248. Lajoie-Mazenc, I. *et al.* MAP1A Light Chain-2 Interacts with GTP-RhoB to Control Epidermal Growth Factor (EGF)-dependent EGF Receptor Signaling. *J. Biol. Chem.* **283**, 4155–4164 (2008).

249. Gutierrez, E. *et al.* Regulation of RhoB Gene Expression during Tumorigenesis and Aging Process and Its Potential Applications in These Processes. *Cancers (Basel).* **11**, 818 (2019).
250. Johnson, D. I. & Pringle, J. R. Molecular characterization of CDC42, a Saccharomyces cerevisiae gene involved in the development of cell polarity. *J. Cell Biol.* **111**, 143–52 (1990).
251. Nobes, C. D. & Hall, A. *Rho, Rac, and Cdc42 GTPases Regulate the Assembly of Multimolecular Focal Complexes Associated with Actin Stress Fibers, Lamellipodia, and Filopodia*. *Cell* **81**, (1995).
252. Welch, M. D. & Mullins, R. D. Cellular Control of Actin Nucleation. *Annu. Rev. Cell Dev. Biol.* **18**, 247–288 (2002).
253. Peng, J., Wallar, B. J., Flanders, A., Swiatek, P. J. & Alberts, A. S. Disruption of the Diaphanous-related formin Drf1 gene encoding mDia1 reveals a role for Drf3 as an effector for Cdc42. *Curr. Biol.* **13**, 534–45 (2003).
254. Garvalov, B. K. *et al.* Cdc42 Regulates Cofilin during the Establishment of Neuronal Polarity. *J. Neurosci.* **27**, 13117–13129 (2007).
255. Czuchra, A. *et al.* Cdc42 Is Not Essential for Filopodium Formation, Directed Migration, Cell Polarization, and Mitosis in Fibroblastoid Cells. *Mol. Biol. Cell* **16**, 4473–4484 (2005).
256. Anderson, S. *et al.* MYC-nick promotes cell migration by inducing fascin expression and Cdc42 activation. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5481-90 (2016).
257. Lin, D. *et al.* A mammalian PAR-3–PAR-6 complex implicated in Cdc42/Rac1 and aPKC signalling and cell polarity. *Nat. Cell Biol.* **2**, 540–547 (2000).
258. Feng, Q. *et al.* Cool-1 functions as an essential regulatory node for EGFreceptor- and Src-mediated cell growth. *Nat. Cell Biol.* **8**, 945–956 (2006).
259. Corbetta, S. *et al.* Generation and Characterization of Rac3 Knockout Mice. *Mol. Cell. Biol.* **25**, 5763–5776 (2005).
260. Sugihara, K. *et al.* Rac1 is required for the formation of three germ layers during gastrulation. *Oncogene* **17**, 3427–3433 (1998).
261. Roberts, A. W. *et al.* Deficiency of the hematopoietic cell-specific Rho family GTPase Rac2 is characterized by abnormalities in neutrophil function and host defense. *Immunity* **10**, 183–96 (1999).
262. Payapilly, A. & Malliri, A. Compartmentalisation of RAC1 signalling. *Curr. Opin. Cell Biol.* **54**, 50–56 (2018).
263. Michaelson, D. *et al.* Differential Localization of Rho Gtpases in Live Cells. *J. Cell Biol.* **152**, 111–126 (2001).
264. Lanning, C. C., Daddona, J. L., Ruiz-Velasco, R., Shafer, S. H. & Williams, C. L. The Rac1 C-terminal Polybasic Region Regulates the Nuclear Localization and Protein Degradation of Rac1. *J Biol Chem*. **279**, 44197-210 (2004).
265. Michaelson, D. *et al.* Rac1 accumulates in the nucleus during the G2 phase of the cell cycle and promotes cell division. *J. Cell Biol.* **181**, 485–496 (2008).
266. Ridley, A. J., Paterson, H. F., Johnston, C. L., Diekmann, D. & Hall, A. The small GTP-binding protein rac regulates growth factor-induced membrane ruffling. *Cell* **70**, 401–410 (1992).
267. Ridley, A. J. Rho GTPase signalling in cell migration. *Curr. Opin. Cell Biol.* **36**, 103–112 (2015).
268. Lawson, C. D. & Burridge, K. The on-off relationship of Rho and Rac during integrin-mediated adhesion and cell migration. *Small GTPases* **5**, e27958 (2014).
269. Jaffe, A. B. & Hall, A. RHO GTPASES: Biochemistry and Biology. *Annu. Rev. Cell Dev. Biol.* **21**, 247–269 (2005).
270. Machacek, M. *et al.* Coordination of Rho GTPase activities during cell protrusion. *Nature* **461**, 99–103 (2009).
271. Wu, C. *et al.* Arp2/3 Is Critical for Lamellipodia and Response to Extracellular Matrix Cues but Is Dispensable for Chemotaxis. *Cell* **148**, 973–987 (2012).
272. Veltman, D. M., King, J. S., Machesky, L. M. & Insall, R. H. SCAR knockouts in Dictyostelium: WASP assumes SCAR's position and upstream regulators in pseudopods. *J. Cell Biol.* **198**, 501–8 (2012).
273. Vasioukhin, V., Bauer, C., Yin, M. & Fuchs, E. Directed actin polymerization is the driving force for epithelial cell-cell adhesion. *Cell* **100**, 209–19 (2000).

274. Ehrlich, J. S., Hansen, M. D. H. & Nelson, W. J. Spatio-temporal regulation of Rac1 localization and lamellipodia dynamics during epithelial cell-cell adhesion. *Dev. Cell* **3**, 259–70 (2002).
275. Giannone, G. *et al.* Lamellipodial actin mechanically links myosin activity with adhesion-site formation. *Cell* **128**, 561–75 (2007).
276. Pasapera, A. M. *et al.* Rac1-Dependent Phosphorylation and Focal Adhesion Recruitment of Myosin IIA Regulates Migration and Mechanosensing. *Curr. Biol.* **25**, 175–186 (2015).
277. Yamada, S. & Nelson, W. J. Localized zones of Rho and Rac activities drive initiation and expansion of epithelial cell-cell adhesion. *J. Cell Biol.* **178**, 517–27 (2007).
278. Chang, F., Lemmon, C. A., Park, D. & Romer, L. H. FAK Potentiates Rac1 Activation and Localization to Matrix Adhesion Sites: A Role for βPIX. *Mol. Biol. Cell* **18**, 253–264 (2007).
279. Carmon, K. S. *et al.* LGR5 receptor promotes cell-cell adhesion in stem cells and colon cancer cells via the IQGAP1-Rac1 pathway. *J Biol Chem.* **292**, 14989-15001 (2017).
280. Malliri, A., van Es, S., Huveneers, S. & Collard, J. G. The Rac Exchange Factor Tiam1 Is Required for the Establishment and Maintenance of Cadherin-based Adhesions. *J. Biol. Chem.* **279**, 30092–30098 (2004).
281. Nelson, W. J. Regulation of cell-cell adhesion by the cadherin-catenin complex. *Biochem. Soc. Trans.* **36**, 149–55 (2008).
282. Yoshida, S., Hoppe, A. D., Araki, N. & Swanson, J. A. Sequential signaling in plasma-membrane domains during macropinosome formation in macrophages. *J. Cell Sci.* **122**, 3250–3261 (2009).
283. Wells, C. M., Walmsley, M., Ooi, S., Tybulewicz, V. & Ridley, A. J. Rac1-deficient macrophages exhibit defects in cell spreading and membrane ruffling but not migration. *J. Cell Sci.* **117**, 1259–1268 (2004).
284. Ikeda, Y. *et al.* Rac1 switching at the right time and location is essential for Fcγ receptor-mediated phagosome formation. *J. Cell Sci.* **130**, 2530–2540 (2017).
285. Abo, A. *et al.* Activation of the NADPH oxidase involves the small GTP-binding protein p21rac1. *Nature* **353**, 668–670 (1991).
286. den Hartog, G. *et al.* Regulation of Rac1 and Reactive Oxygen Species Production in Response to Infection of Gastrointestinal Epithelia. *PLOS Pathog.* **12**, e1005382 (2016).
287. Debidda, M., Williams, D. A. & Zheng, Y. Rac1 GTPase regulates cell genomic stability and senescence. *J. Biol. Chem.* **281**, 38519–28 (2006).
288. Nimnual, A. S., Taylor, L. J. & Bar-Sagi, D. Redox-dependent downregulation of Rho by Rac. *Nat. Cell Biol.* **5**, 236–241 (2003).
289. Rygiel, T. P., Mertens, A. E., Strumane, K., van der Kammen, R. & Collard, J. G. The Rac activator Tiam1 prevents keratinocyte apoptosis by controlling ROS-mediated ERK phosphorylation. *J. Cell Sci.* **121**, 1183–92 (2008).
290. Le Belle, J. E. *et al.* Proliferative Neural Stem Cells Have High Endogenous ROS Levels that Regulate Self-Renewal and Neurogenesis in a PI3K/Akt-Dependant Manner. *Cell Stem Cell* **8**, 59–71 (2011).
291. Ito, K. *et al.* Reactive oxygen species act through p38 MAPK to limit the lifespan of hematopoietic stem cells. *Nat. Med.* **12**, 446–451 (2006).
292. Myant, K. B. *et al.* ROS production and NF-κB activation triggered by RAC1 facilitate WNT-driven intestinal stem cell proliferation and colorectal cancer initiation. *Cell Stem Cell* **12**, 761–73 (2013).
293. Coso, O. A. *et al.* The small GTP-binding proteins Rac1 and Cdc42 regulate the activity of the JNK/SAPK signaling pathway. *Cell* **81**, 1137–46 (1995).
294. Zhang, S. *et al.* Rho family GTPases regulate p38 mitogen-activated protein kinase through the downstream mediator Pak1. *J. Biol. Chem.* **270**, 23934–6 (1995).
295. Eblen, S. T., Slack, J. K., Weber, M. J. & Catling, A. D. Rac-PAK signaling stimulates extracellular signal-regulated kinase (ERK) activation by regulating formation of MEK1-ERK complexes. *Mol. Cell. Biol.* **22**, 6023–33 (2002).
296. Slack-Davis, J. K. *et al.* PAK1 phosphorylation of MEK1 regulates fibronectin-stimulated MAPK activation. *J. Cell Biol.* **162**, 281–291 (2003).
297. Saarikangas, J., Zhao, H. & Lappalainen, P. Regulation of the Actin Cytoskeleton-Plasma

Membrane Interplay by Phosphoinositides. *Physiol. Rev.* **90**, 259–289 (2010).
298. Tolias, K. F., Cantley, L. C. & Carpenter, C. L. Rho family GTPases bind to phosphoinositide kinases. *J. Biol. Chem.* **270**, 17656–9 (1995).
299. Hemmings, B. A. & Restuccia, D. F. PI3K-PKB/Akt Pathway. *Cold Spring Harb. Perspect. Biol.* **4**, a011189–a011189 (2012).
300. Murga, C., Zohar, M., Teramoto, H. & Gutkind, J. S. Rac1 and RhoG promote cell survival by the activation of PI3K and Akt, independently of their ability to stimulate JNK and NF-κB. *Oncogene* **21**, 207–216 (2002).
301. Zhu, G. *et al.* An EGFR/PI3K/AKT axis promotes accumulation of the Rac1-GEF Tiam1 that is critical in EGFR-driven tumorigenesis. *Oncogene* **34**, 5971–5982 (2015).
302. Saci, A., Cantley, L. C. & Carpenter, C. L. Rac1 Regulates the Activity of mTORC1 and mTORC2 and Controls Cellular Size. *Mol. Cell* **42**, 50–61 (2011).
303. Morrison Joly, M. *et al.* Two distinct mTORC2-dependent pathways converge on Rac1 to drive breast cancer metastasis. *Breast Cancer Res.* **19**, 74 (2017).
304. Olson, M., Ashworth, A. & Hall, A. An essential role for Rho, Rac, and Cdc42 GTPases in cell cycle progression through G1. *Science.* **269**, 1270–1272 (1995).
305. Maroto, B., Ye, M. B., von Lohneysen, K., Schnelzer, A. & Knaus, U. G. P21-activated kinase is required for mitotic progression and regulates Plk1. *Oncogene* **27**, 4900–4908 (2008).
306. Schmidt, A., Durgan, J., Magalhaes, A. & Hall, A. Rho GTPases regulate PRK2/PKN2 to control entry into mitosis and exit from cytokinesis. *EMBO J.* **26**, 1624–1636 (2007).
307. May, M., Schelle, I., Brakebusch, C., Rottner, K. & Genth, H. Rac1-dependent recruitment of PAK2 to G2 phase centrosomes and their roles in the regulation of mitotic entry. *Cell Cycle* **13**, 2210–2220 (2014).
308. Mercurio, F. & Manning, A. M. Multiple signals converging on NF-κB. *Curr. Opin. Cell Biol.* **11**, 226–232 (1999).
309. Sulciner, D. J. *et al.* rac1 regulates a cytokine-stimulated, redox-dependent pathway necessary for NF-kappaB activation. *Mol. Cell. Biol.* **16**, 7115–21 (1996).
310. Perona, R. *et al.* Activation of the nuclear factor-kappaB by Rho, CDC42, and Rac-1 proteins. *Genes Dev.* **11**, 463–475 (1997).
311. Boyer, L. *et al.* Rac GTPase Instructs Nuclear Factor-κB Activation by Conveying the SCF Complex and IkBα to the Ruffling Membranes. *Mol. Biol. Cell* **15**, 1124–1133 (2004).
312. Arbibe, L. *et al.* Toll-like receptor 2–mediated NF-κB activation requires a Rac1-dependent pathway. *Nat. Immunol.* **1**, 533–540 (2000).
313. Kim, S. J. & Yoon, S. Activated Rac1 regulates the degradation of IκBα and the nuclear translocation of STAT3–NFκB complexes in starved cancer cells. *Exp. Mol. Med.* **48**, e231–e231 (2016).
314. Esufali, S. & Bapat, B. Cross-talk between Rac1 GTPase and dysregulated Wnt signaling pathway leads to cellular redistribution of b-catenin and TCF/LEF-mediated transcriptional activation. *Oncogene* **23**, 8260–8271 (2004).
315. Jamieson, C., Lui, C., Brocardo, M. G., Martino-Echarri, E. & Henderson, B. R. Rac1 augments Wnt signaling by stimulating β-catenin-lymphoid enhancer factor-1 complex assembly independent of β-catenin nuclear import. *J. Cell Sci.* **128**, 3933–46 (2015).
316. Wu, X. *et al.* Rac1 activation controls nuclear localization of beta-catenin during canonical Wnt signaling. *Cell* **133**, 340–53 (2008).
317. Phelps, R. A. *et al.* A Two-Step Model for Colon Adenoma Initiation and Progression Caused by APC Loss. *Cell* **137**, 623–634
318. Bosco, E. E., Nakai, Y., Hennigan, R. F., Ratner, N. & Zheng, Y. NF2-deficient cells depend on the Rac1-canonical Wnt signaling pathway to promote the loss of contact inhibition of proliferation. *Oncogene* **29**, 2540–2549 (2010).
319. Fanto, M., Weber, U., Strutt, D. I. & Mlodzik, M. Nuclear signaling by Rac and Rho GTPases is required in the establishment of epithelial planar polarity in the Drosophila eye. *Curr. Biol.* **10**, 979–88 (2000).
320. Gómez-Orte, E., Sáenz-Narciso, B., Moreno, S. & Cabello, J. Multiple functions of the noncanonical Wnt pathway. *Trends Genet.* **29**, 545–553 (2013).

321. Qiu, R.-G., Chen, J., Kirn, D., McCormick, F. & Symons, M. An essential role for Rac in Ras transformation. *Nature* **374**, 457–459 (1995).
322. Khosravi-Far, R., Solski, P. A., Clark, G. J., Kinch, M. S. & Der, C. J. Activation of Rac1, RhoA, and mitogen-activated protein kinases is required for Ras transformation. *Mol. Cell. Biol.* **15**, 6443–53 (1995).
323. Kissil, J. L. *et al.* Requirement for Rac1 in a K-ras–Induced Lung Cancer in the Mouse. *Cancer Res.* **67**, 8089–8094 (2007).
324. Hodis, E. *et al.* A Landscape of Driver Mutations in Melanoma. *Cell* **150**, 251–263 (2012).
325. Krauthammer, M. *et al.* Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nat. Genet.* **44**, 1006–1014 (2012).
326. Watson, I. R. *et al.* The RAC1 P29S Hotspot Mutation in Melanoma Confers Resistance to Pharmacological Inhibition of RAF. *Cancer Res.* **74**, 4845–4852 (2014).
327. Chang, M. T. *et al.* Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat. Biotechnol.* **34**, 155–163 (2016).
328. Andrio, E. *et al.* Identification of cancer-associated missense mutations in hace1 that impair cell growth control and Rac1 ubiquitylation. *Sci. Rep.* **7**, 44779 (2017).
329. Goka, E. T. & Lippman, M. E. Loss of the E3 ubiquitin ligase HACE1 results in enhanced Rac1 signaling contributing to breast cancer progression. *Oncogene* **34**, 5395–5405 (2015).
330. Katzav, S., Martin-Zanca, D. & Barbacid, M. vav, a novel human oncogene derived from a locus ubiquitously expressed in hematopoietic cells. *EMBO J.* **8**, 2283–90 (1989).
331. Miki, T., Smith, C. L., Long, J. E., Eva, A. & Fleming, T. P. Oncogene ect2 is related to regulators of small GTP-binding proteins. *Nature* **362**, 462–465 (1993).
332. Habets, G. G. *et al.* Identification of an invasion-inducing gene, Tiam-1, that encodes a protein with homology to GDP-GTP exchangers for Rho-like proteins. *Cell* **77**, 537–49 (1994).
333. Cook, D. R., Rossman, K. L. & Der, C. J. Rho guanine nucleotide exchange factors: regulators of Rho GTPase activity in development and disease. *Oncogene* **33**, 4021–4035 (2014).
334. Sosa, M. S. *et al.* Identification of the Rac-GEF P-Rex1 as an Essential Mediator of ErbB Signaling in Breast Cancer. *Mol. Cell* **40**, 877–892 (2010).
335. Lindsay, C. R. *et al.* P-Rex1 is required for efficient melanoblast migration and melanoma metastasis. *Nat. Commun.* **2**, 555 (2011).
336. Qin, J. *et al.* Upregulation of PIP3-dependent Rac exchanger 1 (P-Rex1) promotes prostate cancer metastasis. *Oncogene* **28**, 1853–1863 (2009).
337. Goel, H. L. *et al.* P-Rex1 Promotes Resistance to VEGF/VEGFR-Targeted Therapy in Prostate Cancer. *Cell Rep.* **14**, 2193–2208 (2016).
338. Kim, M., Nozu, F., Kusama, K. & Imawari, M. Cholecystokinin stimulates the recruitment of the Src–RhoA–phosphoinositide 3-kinase pathway by Vav-2 downstream of Gα13 in pancreatic acini. *Biochem. Biophys. Res. Commun.* **339**, 271–276 (2006).
339. Chiariello, M., Marinissen, M. J. & Gutkind, J. S. Regulation of c-myc expression by PDGF through Rho GTPases. *Nat. Cell Biol.* **3**, 580–586 (2001).
340. Bartolomé, R. A. *et al.* Activation of Vav/Rho GTPase Signaling by CXCL12 Controls Membrane-Type Matrix Metalloproteinase–Dependent Melanoma Cell Invasion. *Cancer Res.* **66**, 248–258 (2006).
341. Patel, V. *et al.* Persistent activation of Rac1 in squamous carcinomas of the head and neck: evidence for an EGFR/Vav2 signaling axis involved in cell invasion. *Carcinogenesis* **28**, 1145–1152 (2007).
342. Lambert, J. M. *et al.* Tiam1 mediates Ras activation of Rac by a PI(3)K-independent mechanism. *Nat. Cell Biol.* **4**, 621–625 (2002).
343. Porter, A. P., Papaioannou, A. & Malliri, A. Deregulation of Rho GTPases in cancer. *Small GTPases* **7**, 123–138 (2016).
344. Minard, M. E., Ellis, L. M. & Gallick, G. E. Tiam1 regulates cell adhesion, migration and apoptosis in colon tumor cells. *Clin. Exp. Metastasis* **23**, 301–313 (2006).
345. Malliri, A. *et al.* The rac activator Tiam1 is a Wnt-responsive gene that modifies intestinal tumor development. *J. Biol. Chem.* **281**, 543–8 (2006).
346. Malliri, A. *et al.* Mice deficient in the Rac activator Tiam1 are resistant to Ras-induced skin

tumours. *Nature* **417**, 867–871 (2002).
347. Stebel, A., Brachetti, C., Kunkel, M., Schmidt, M. & Fritz, G. Progression of breast tumors is accompanied by a decrease in expression of the Rho guanine exchange factor Tiam1. *Oncol. Rep.* **21**, 217–22 (2009).
348. Diamantopoulou, Z. *et al.* TIAM1 Antagonizes TAZ/YAP Both in the Destruction Complex in the Cytoplasm and in the Nucleus to Inhibit Invasion of Intestinal Epithelial Cells. *Cancer Cell* **31**, 621–634.e6 (2017).
349. Laurin, M. *et al.* Rac-specific guanine nucleotide exchange factor DOCK1 is a critical regulator of HER2-mediated breast cancer metastasis. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7434–9 (2013).
350. Hirotada Tajiri, A. *et al.* Targeting Ras-Driven Cancer Cell Survival and Invasion through Selective Inhibition of DOCK1. *Cell Rep.* **19**, 969–980 (2017).
351. Lahoz, A. & Hall, A. A tumor suppressor role for srGAP3 in mammary epithelial cells. *Oncogene* **32**, 4854–4860 (2013).
352. Marko, T. A. *et al.* Slit-Robo GTPase-Activating Protein 2 as a metastasis suppressor in osteosarcoma. *Sci. Rep.* **6**, 39059 (2016).
353. Yang, C., Liu, Y., Lemmon, M. A. & Kazanietz, M. G. Essential role for Rac in heregulin beta1 mitogenic signaling: a mechanism that involves epidermal growth factor receptor and is independent of ErbB4. *Mol. Cell. Biol.* **26**, 831–42 (2006).
354. Imaoka, H. *et al.* RacGAP1 expression, increasing tumor malignant potential, as a predictive biomarker for lymph node metastasis and poor prognosis in colorectal cancer. *Carcinogenesis* **36**, 346–354 (2015).
355. Mi, S. *et al.* RNA-seq Identification of RACGAP1 as a Metastatic Driver in Uterine Carcinosarcoma. *Clin. Cancer Res.* **22**, 4676–4686 (2016).
356. He, Y. *et al.* The Cdc42/Rac1 regulator CdGAP is a novel E-cadherin transcriptional co-repressor with Zeb2 in breast cancer. *Oncogene* **36**, 3490–3503 (2017).
357. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
358. Yang, X. *et al.* Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell* **164**, 805–817 (2016).
359. Wang, Z. *et al.* Systematic identification and analysis of exonic splicing silencers. *Cell* **119**, 831–45 (2004).
360. Crawford, J. B. & Patton, J. G. Activation of -Tropomyosin Exon 2 Is Regulated by the SR Protein 9G8 and Heterogeneous Nuclear Ribonucleoproteins H and F. *Mol. Cell. Biol.* **26**, 8791–8802 (2006).
361. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
362. Weischenfeldt, J. *et al.* Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. *Genome Biol.* **13**, R35 (2012).
363. Gonzàlez-Porta, M., Frankish, A., Rung, J., Harrow, J. & Brazma, A. Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. *Genome Biol.* **14**, R70 (2013).
364. Singh, R. K. & Cooper, T. A. Pre-mRNA splicing in disease and therapeutics. *Trends Mol. Med.* **18**, 472–82 (2012).
365. Sveen, A., Kilpinen, S., Ruusulehto, A., Lothe, R. A. & Skotheim, R. I. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* **35**, 2413–2427 (2016).
366. Boise, L. H. *et al.* bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell* **74**, 597–608 (1993).
367. Bae, J., Leo, C. P., Hsu, S. Y. & Hsueh, A. J. W. MCL-1S, a Splicing Variant of the Antiapoptotic BCL-2 Family Member MCL-1, Encodes a Proapoptotic Protein Possessing Only the BH3 Domain. *J. Biol. Chem.* **275**, 25255–25261 (2000).
368. Ladomery, M. R., Harper, S. J. & Bates, D. O. Alternative splicing in angiogenesis: The vascular

endothelial growth factor paradigm. *Cancer Lett.* **249**, 133–142 (2007).
369. Orian-Rousseau, V. CD44 Acts as a Signaling Platform Controlling Tumor Progression and Metastasis. *Front. Immunol.* **6**, 154 (2015).
370. Michael Seiler, S. P., Anant A. Agrawal, T. C. G., Atlas Research Network, *et al.* Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep.* **23**, 282–296 (2018).
371. Kahles, A. *et al.* Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell* **34**, 211–224.e6 (2018).
372. Jordan, P., BrazÃ£o, R., Boavida, M. G., Gespach, C. & Chastre, E. Cloning of a novel human Rac1b splice variant with increased expression in colorectal tumors. *Oncogene* **18**, 6835–6839 (1999).
373. Schnelzer, A. *et al.* Rac1 in human breast cancer: overexpression, mutation analysis, and characterization of a new isoform, Rac1b. *Oncogene*. **19**, 3013-3020 (2000)-
374. Matos, P. *et al.* Small GTPase Rac1: Structure, Localization, and Expression of the Human Gene. *Biochem. Biophys. Res. Commun.* **277**, 741–751 (2000).
375. Matos, P., Collard, J. G. & Jordan, P. Tumor-related Alternatively Spliced Rac1b Is Not Regulated by Rho-GDP Dissociation Inhibitors and Exhibits Selective Downstream Signaling. *J Biol Chem*. **278**, 50442-8 (2003).
376. Fiegen, D. *et al.* Alternative splicing of Rac1 generates Rac1b, a self-activating GTPase. *J. Biol. Chem.* **279**, 4743–9 (2004).
377. Singh, A. *et al.* Rac1b, a tumor associated, constitutively active Rac1 splice variant, promotes cellular transformation. *Oncogene* **23**, 9369–9380 (2004).
378. Orlichenko, L. *et al.* The 19-Amino Acid Insertion in the Tumor-associated Splice Isoform Rac1b Confers Specific Binding to p120 Catenin. *J. Biol. Chem.* **285**, 19153–19161 (2010).
379. Ying, L. *et al.* Sanguinarine inhibits Rac1b-rendered cell survival enhancement by promoting apoptosis and blocking proliferation. *Acta Pharmacol. Sin.* **36**, 229–240 (2015).
380. Li, G. *et al.* Rac1b enhances cell survival through activation of the JNK2/c-JUN/Cyclin-D1 and AKT2/MCL1 pathways. *Oncotarget* **7**, 17970–17985 (2016).
381. Matos, P. & Jordan, P. Expression of Rac1b stimulates NF-κB-mediated cell survival and G1/S progression. *Exp. Cell Res.* **305**, 292–299 (2005).
382. Matos, P. & Jordan, P. Rac1, but Not Rac1B, Stimulates RelB-mediated Gene Transcription in Colorectal Cancer Cells. *J. Biol. Chem.* **281**, 13724–13732 (2006).
383. Matos, P. & Jordan, P. Increased Rac1b Expression Sustains Colorectal Tumor Cell Survival. *Mol. Cancer Res.* **6**, 1178–1184 (2008).
384. Faria, M., Capinha, L., Simões-Pereira, J., Bugalho, M. J. & Silva, A. L. Extending the Impact of RAC1b Overexpression to Follicular Thyroid Carcinomas. *Int. J. Endocrinol.* **2016**, 1–6 (2016).
385. Faria, M. *et al.* RAC1b overexpression stimulates proliferation and NF-kB-mediated anti-apoptotic signaling in thyroid cancer cells. *PLoS One* **12**, e0172689 (2017).
386. Esufali, S., Charames, G. S., Pethe, V. V, Buongiorno, P. & Bapat, B. Activation of tumor-specific splice variant Rac1b by dishevelled promotes canonical Wnt signaling and decreased adhesion of colorectal cancer cells. *Cancer Res.* **67**, 2469–79 (2007).
387. Bapat, B., Charames, G. S. & Bapat, B. Rac1b recruits Dishevelled and β-catenin to Wnt target gene promoters independent of Wnt3A stimulation. *Int. J. Oncol.* **39**, 805–810 (2011).
388. Fischer, K. R. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472–476 (2015).
389. Pelisch, F. *et al.* Involvement of hnRNP A1 in the matrix metalloprotease-3-dependent regulation of Rac1 pre-mRNA splicing. *J. Cell. Biochem.* **113**, 2319–2329 (2012).
390. Radisky, D. C. *et al.* Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability. *Nature* **436**, 123–127 (2005).
391. Mehner, C. *et al.* Tumor Cell–Derived MMP3 Orchestrates Rac1b and Tissue Alterations That Promote Pancreatic Adenocarcinoma. *Mol. Cancer Res.* **12**, 1430–1439 (2014).
392. Nelson, C. M., Khauv, D., Bissell, M. J. & Radisky, D. C. Change in cell shape is required for matrix metalloproteinase-induced epithelial-mesenchymal transition of mammary epithelial cells. *J. Cell. Biochem.* **105**, 25–33 (2008).

393. Lee, K. *et al.* Matrix compliance regulates Rac1b localization, NADPH oxidase assembly, and epithelial-mesenchymal transition. *Mol. Biol. Cell* **23**, 4097–4108 (2012).
394. Chen, Q. K., Lee, K., Radisky, D. C. & Nelson, C. M. Extracellular matrix proteins regulate epithelial–mesenchymal transition in mammary epithelial cells. *Differentiation* **86**, 126–132 (2013).
395. Stallings-Mann, M. L. *et al.* Matrix metalloproteinase induction of Rac1b, a key effector of lung cancer progression. *Sci. Transl. Med.* **4**, 142ra95 (2012).
396. Dai, B. *et al.* Blockade of ARHGAP11A reverses malignant progress via inactivating Rac1B in hepatocellular carcinoma. *Cell Commun. Signal.* **16**, 99 (2018).
397. Nimnual, A. S., Taylor, L. J., Nyako, M., Jeng, H.-H. & Bar-Sagi, D. Perturbation of cytoskeleton dynamics by the opposing effects of Rac1 and Rac1b. *Small GTPases* **1**, 89–97 (2010).
398. Wang, F. *et al.* SPSB1-mediated HnRNP A1 ubiquitylation regulates alternative splicing and cell migration in EGF signaling. *Cell Res.* **27**, 540–558 (2017).
399. Xing, S. *et al.* DIS3L2 promotes progression of hepatocellular carcinoma via hnRNP U-mediated alternative splicing. *Cancer Res.* **79**, (2019).
400. Ishii, H. *et al.* Epithelial splicing regulatory proteins 1 (ESRP1) and 2 (ESRP2) suppress cancer cell motility via different mechanisms. *J. Biol. Chem.* **289**, 27386–99 (2014).
401. Ungefroren, H. *et al.* Rac1b negatively regulates TGF-B-induced cell motility in pancreatic ductal epithelial cells by suppressing Smad signalling. *Oncotarget* **5**, (2014).
402. Melzer, C., von der Ohe, J., Hass, R. & Ungefroren, H. TGF-β-Dependent Growth Arrest and Cell Migration in Benign and Malignant Breast Epithelial Cells Are Antagonistically Controlled by Rac1 and Rac1b. *Int. J. Mol. Sci.* **18**, 1574 (2017).
403. Witte, D. *et al.* Negative regulation of TGF-β1-induced MKK6-p38 and MEK-ERK signalling and epithelial-mesenchymal transition by Rac1b. *Sci. Rep.* **7**, 17313 (2017).
404. Otterbein, H. *et al.* RAC1B Suppresses TGF-β-Dependent Chemokinesis and Growth Inhibition through an Autoregulatory Feed-Forward Loop Involving PAR2 and ALK5. *Cancers (Basel).* **11**, 1211 (2019).
405. Ungefroren, H. *et al.* RAC1B Suppresses TGF-β1-Dependent Cell Migration in Pancreatic Carcinoma Cells through Inhibition of the TGF-β Type I Receptor ALK5. *Cancers (Basel).* **11**, 691 (2019).
406. Matos, P. *et al.* Ibuprofen Inhibits Colitis-Induced Overexpression of TumorRelated Rac1b. *Neoplasia* **15**, 102–111 (2013).
407. Matos, P. *et al.* B-RafV600E Cooperates With Alternative Spliced Rac1b to Sustain Colorectal Cancer Cell Survival. *Gastroenterology* **135**, 899–906 (2008).
408. Zhou, C. *et al.* The Rac1 splice form Rac1b promotes K-ras-induced lung tumorigenesis. *Oncogene* **32**, 903–909 (2013).
409. Kotelevets, L. *et al.* The Rac1 splice form Rac1b favors mouse colonic mucosa regeneration and contributes to intestinal cancer progression. *Oncogene* **37**, 6054–6068 (2018).
410. Karni, R. *et al.* The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.* **14**, 185–193 (2007).
411. van der Houven van Oordt, W. *et al.* The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. *J. Cell Biol.* **149**, 307–16 (2000).
412. Gonçalves, V., Matos, P. & Jordan, P. Antagonistic SR proteins regulate alternative splicing of tumor-related Rac1b downstream of the PI3-kinase and Wnt pathways. *Hum. Mol. Genet.* **18**, 3696–3707 (2009).
413. Goncalves, V. *et al.* Phosphorylation of SRSF1 by SRPK1 regulates alternative splicing of tumor-related Rac1b in colorectal cells. *RNA* **20**, 474–482 (2014).
414. Sansom, O. J. *et al.* APC short term reference - Loss of Apc in vivo immediately perturbs Wnt signaling, differentiation, and migration. *Genes Dev.* **18**, 1385–90 (2004).
415. Su, L. K. *et al.* Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science* **256**, 668–70 (1992).
416. Moser, A. R. *et al. ApcMin, a mutation in the murine Apc gene, predisposes to mammary carcinomas and focal alveolar hyperplasias*. *Proc. Natl. Acad. Sci. USA* **90**, (1993).

417. Moser, A. R., Pitot, H. C. & Dove, W. F. A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science* **247**, 322–4 (1990).
418. Pollard, P. *et al.* The Apc1322T Mouse Develops Severe Polyposis Associated With Submaximal Nuclear β-Catenin Expression. *Gastroenterology* **136**, 2204–2213.e13 (2009).
419. Robanus-Maandag, E. C. *et al.* A new conditional Apc-mutant mouse model for colorectal cancer. *Carcinogenesis* **31**, 946–952 (2010).
420. Oshima, M. *et al.* Loss of Apc heterozygosity and abnormal tissue building in nascent intestinal polyps in mice carrying a truncated Apc gene. *Proc. Natl. Acad. Sci.* **92**, 4482–4486 (1995).
421. Wei, K., Kucherlapati, R. & Edelmann, W. Mouse models for human DNA mismatch-repair gene defects. *Trends Mol. Med.* **8**, 346–353 (2002).
422. Sodir, N. M. *et al.* Smad3 Deficiency Promotes Tumorigenesis in the Distal Colon of *Apc* $^{Min/+}$ Mice. *Cancer Res.* **66**, 8430–8438 (2006).
423. Halberg, R. B. *et al.* Tumorigenesis in the multiple intestinal neoplasia mouse: Redundancy of negative regulators and specificity of modifiers. *Proc. Natl. Acad. Sci.* **97**, 3461–3466 (2000).
424. Batlle, E. *et al.* EphB receptor activity suppresses colorectal cancer progression. *Nature* **435**, 1126–1130 (2005).
425. Nagy, A. Cre recombinase: The universal reagent for genome tailoring. *Genesis* **26**, 99–109 (2000).
426. Rossant, J. & McMahon, A. "Cre"-ating mouse mutants-a meeting review on conditional mouse genetics. *Genes Dev.* **13**, 142–5 (1999).
427. El Marjou, F. *et al.* Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *genesis* **39**, 186–193 (2004).
428. Tetteh, P. W. *et al.* Generation of an inducible colon-specific Cre enzyme mouse line for colon cancer research. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11859–11864 (2016).
429. Xue, Y., Johnson, R., Desmet, M., Snyder, P. W. & Fleet, J. C. Generation of a Transgenic Mouse for Colorectal Cancer Research with Intestinal Cre Expression Limited to the Large Intestine. *Mol Cancer Res*. **8**, 1095-1104 (2010).
430. Feng, Y. *et al.* Sox9 Induction, Ectopic Paneth Cells, and Mitotic Spindle Axis Defects in Mouse Colon Adenomatous Epithelium Arising From Conditional Biallelic Apc Inactivation. *Am. J. Pathol.* **183**, 493–503 (2013).
431. Malaterre, J. *et al.* c-Myb is required for progenitor cell homeostasis in colonic crypts. *Proc. Natl. Acad. Sci.* **104**, 3829–3834 (2007).
432. Hinoi, T. *et al.* Mouse Model of Colonic Adenoma-Carcinoma Progression Based on Somatic Apc Inactivation. *Cancer Res.* **48**, 1159–1161 (2007).
433. Abremski, K. & Hoess, R. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J. Biol. Chem.* **259**, 1509–14 (1984).
434. Danielian, P. S., Muccino, D., Rowitch, D. H., Michael, S. K. & Mcmahon, A. P. Brief Communication. *Curr. Biol.* **8**, 1323–1326 (1323).
435. Hayashi, S. & Mcmahon, A. P. Efficient Recombination in Diverse Tissues by a Tamoxifen-Inducible Form of Cre: A Tool for Temporally Regulated Gene Activation/Inactivation in the *Mouse. Dev Biol*. **244**, 305-18 (2002).
436. Shibata, H. *et al.* Rapid colorectal adenoma formation initiated by conditional targeting of the Apc gene. *Science* **278**, 120–3 (1997).
437. Jonkers, J. *et al.* Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nat. Genet.* **29**, 418–425 (2001).
438. Kitamura, T. *et al.* Inactivation of chemokine (C-C motif) receptor 1 (CCR1) suppresses colon cancer liver metastasis by blocking accumulation of immature myeloid cells in a mouse model. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 13063–8 (2010).
439. Grabinger, T. *et al.* Ex vivo culture of intestinal crypt organoids as a model system for assessing cell death induction in intestinal epithelial cells and enteropathy. *Cell Death Dis.* **5**, e1228–e1228 (2014).
440. Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet.* **45**, 1113-20 (2013)

441. Ryan, M. *et al.* TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. *Nucleic Acids Res.* **44**, (2016).
442. Wang, Q. *et al.* Data Descriptor: Unifying cancer and normal RNA sequencing data from different sources. *Scientific Data*. **5**, 180061, (2018).
443. Fleming, I. N., Elliott, C. M. & Exton, J. H. Differential Translocation of Rho Family GTPases by Lysophosphatidic Acid, Endothelin-1, and Platelet-derived Growth Factor. *J Biol Chem.* **271**, 33067-73. (1996).
444. Merlos-Suárez, A. *et al.* The Intestinal Stem Cell Signature Identifies Colorectal Cancer Stem Cells and Predicts Disease Relapse. *Cell Stem Cell* **8**, 511–524 (2011).
445. van der Flier, L. G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136**, 903–12 (2009).
446. Bettess, M. D. *et al.* c-Myc is required for the formation of intestinal crypts but dispensable for homeostasis of the adult intestinal epithelium. *Mol. Cell. Biol.* **25**, 7868–78 (2005).
447. Sansom, O. J. *et al.* Myc deletion rescues Apc deficiency in the small intestine. *Nature* **446**, 676–679 (2007).
448. Schnelzer, A. *et al.* Rac1 in human breast cancer: overexpression, mutation analysis, and characterization of a new isoform, Rac1b. *Oncogene* **19**, 3013–3020 (2000).
449. Gao, J. *et al.* Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* **6**, pl1-pl1 (2013).
450. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discov.* **2**, 401–404 (2012).
451. UCSC Xena. Available at: http://xena.ucsc.edu/?ref=labworm. (Accessed: 27th May 2019)
452. Ryan, M. C., Cleland, J., Kim, R., Wong, W. C. & Weinstein, J. N. SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. *Bioinformatics* **28**, 2385–2387 (2012).
453. Sun, W. *et al.* TSVdb: a web-tool for TCGA splicing variants analysis. *BMC Genomics* **19**, 405 (2018).
454. Li, J. *et al.* Explore, Visualize, and Analyze Functional Cancer Proteomic Data Using the Cancer Proteome Atlas. *Cancer Res*. **77**, e51-54 (2017).
455. Li, J. *et al.* TCPA: a resource for cancer functional proteomics data. *Nat. Methods* **10**, 1046–1047 (2013).
456. Wang, Y. *et al.* Comprehensive Molecular Characterization of the Hippo Signaling Pathway in Cancer. *Cell Rep.* **25**, 1304–1317.e5 (2018).
457. Ge, Z. *et al.* Integrated Genomic Analysis of the Ubiquitin Pathway across Cancer Types. *Cell Rep.* **23**, 213–226.e3 (2018).
458. Falcon, T. *et al.* Analysis of the Cancer Genome Atlas Data Reveals Novel Putative ncRNAs Targets in Hepatocellular Carcinoma. *Biomed Res. Int.* **2018**, 1–9 (2018).
459. Cava, C., Bertoli, G. & Castiglioni, I. In silico identification of drug target pathways in breast cancer subtypes using pathway cross-talk inhibition. *J. Transl. Med.* **16**, 154 (2018).
460. Colnot, S. *et al.* Colorectal cancers in a new mouse model of familial adenomatous polyposis: influence of genetic and environmental modifiers. *Lab. Investig.* **84**, 1619–1630 (2004).
461. Muñoz, N. M. *et al.* Transforming Growth Factor β Receptor Type II Inactivation Induces the Malignant Transformation of Intestinal Neoplasms Initiated by *Apc* Mutation. *Cancer Res.* **66**, 9837–9844 (2006).
462. Bennecke, M. *et al.* Ink4a/Arf and oncogene-induced senescence prevent tumor progression during alternative colorectal tumorigenesis. *Cancer Cell* **18**, 135–46 (2010).
463. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810 (2012).
464. Kim, D. I. *et al.* Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2453-61 (2014).
465. Roux, K. J., Kim, D. I. & Burke, B. BioID: A Screen for Protein-Protein Interactions. in *Current Protocols in Protein Science* **74**, 19.23.1-19.23.14 (John Wiley & Sons, Inc., 2013).
466. Zhou, G. *et al.* NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene

expression profiling and meta-analysis. *Nucleic Acids Res.* **47**, W234-41 (2019).
467. Xia, J., Gill, E. E. & Hancock, R. E. W. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat. Protoc.* **10**, 823–844 (2015).
468. Davidson, G. *et al.* Cell cycle control of wnt receptor activation. *Dev. Cell* **17**, 788–99 (2009).
469. Katoh, M. Canonical and non-canonical WNT signaling in cancer stem cells and their niches: Cellular heterogeneity, omics reprogramming, targeted therapy and tumor plasticity (Review). *Int. J. Oncol.* **51**, 1357–1369 (2017).
470. Pasquale, E. B. Eph receptors and ephrins in cancer: bidirectional signalling and beyond. *Nat. Rev. Cancer* **10**, 165–180 (2010).
471. Linggi, B. & Carpenter, G. ErbB receptors: new insights on mechanisms and biology. *Trends in Cell Biology*, **16**, 649-656.
472. de Bono, J. S. & Rowinsky, E. K. The ErbB receptor family: a therapeutic target for cancer. *Trends Mol. Med.* **8**, S19-26 (2002).
473. Fürstenberger, G. & Senn, H.-J. Insulin-like growth factors and cancer. *Lancet. Oncol.* **3**, 298–302 (2002).
474. Fan, P.-D. *et al.* YES1 amplification is a mechanism of acquired resistance to EGFR inhibitors identified by transposon mutagenesis and clinical genomics. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E6030–E6038 (2018).
475. Ranganathan, S. *et al.* Loss of EGFR-ASAP1 signaling in metastatic and unresectable hepatoblastoma. *Scientific Reports*. **6**, 38347. (2016).
476. Yoon, H.-Y., Lee, J.-S. & Randazzo, P. A. ARAP1 regulates endocytosis of EGFR. *Traffic* **9**, 2236–52 (2008).
477. Tarcic, G. *et al.* An Unbiased Screen Identifies the DEP-1 Tumour Suppressor as a Phosphatase Controlling EGFR Endocytosis. *Curr. Biol.* **19**, 1788 (2009).
478. Jeong, W.-J. *et al.* Ras Stabilization Through Aberrant Activation of Wnt/ -Catenin Signaling Promotes Intestinal Tumorigenesis. *Sci. Signal.* **5**, ra30-ra30 (2012).
479. Grainger, S. *et al.* EGFR is required for Wnt9a–Fzd9b signalling specificity in haematopoietic stem cells. *Nat. Cell Biol.* **21**, 721-730 (2019).
480. Hynes, N. E. & Lane, H. A. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat. Rev. Cancer* **5**, 341–354 (2005).
481. Burgess, A. W. *et al.* An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol. Cell* **12**, 541–52 (2003).
482. Wieduwilt, M. J. & Moasser, M. M. The epidermal growth factor receptor family: Biology driving targeted therapeutics. *Cell Mol Life Sci*. **65**, 1566-84 (2008).
483. Graus-Porta, D., Beerli, R. R., Daly, J. M. & Hynes, N. E. ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *EMBO J.* **16**, 1647–55 (1997).
484. Yarden, Y. & Sliwkowski, M. X. Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol.* **2**, 127–137 (2001).
485. Normanno, N. *et al.* Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene* **366**, 2–16 (2006).
486. Network, T. C. G. A. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
487. Mendelsohn, J. & Baselga, J. Status of epidermal growth factor receptor antagonists in the biology and treatment of cancer. *J. Clin. Oncol.* **21**, 2787–99 (2003).
488. Madshus, I. H. & Stang, E. Internalization and intracellular sorting of the EGF receptor: a model for understanding the mechanisms of receptor trafficking. *J. Cell Sci.* **122**, 3433–3439 (2009).
489. Kinali, M. *et al.* Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study. *Lancet Neurol.* **8**, 918–928 (2009).
490. Hua, Y. *et al.* Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes Dev.* **24**, 1634–1644 (2010).
491. Sato, T. *et al.* Long-term Expansion of Epithelial Organoids From Human Colon, Adenoma,

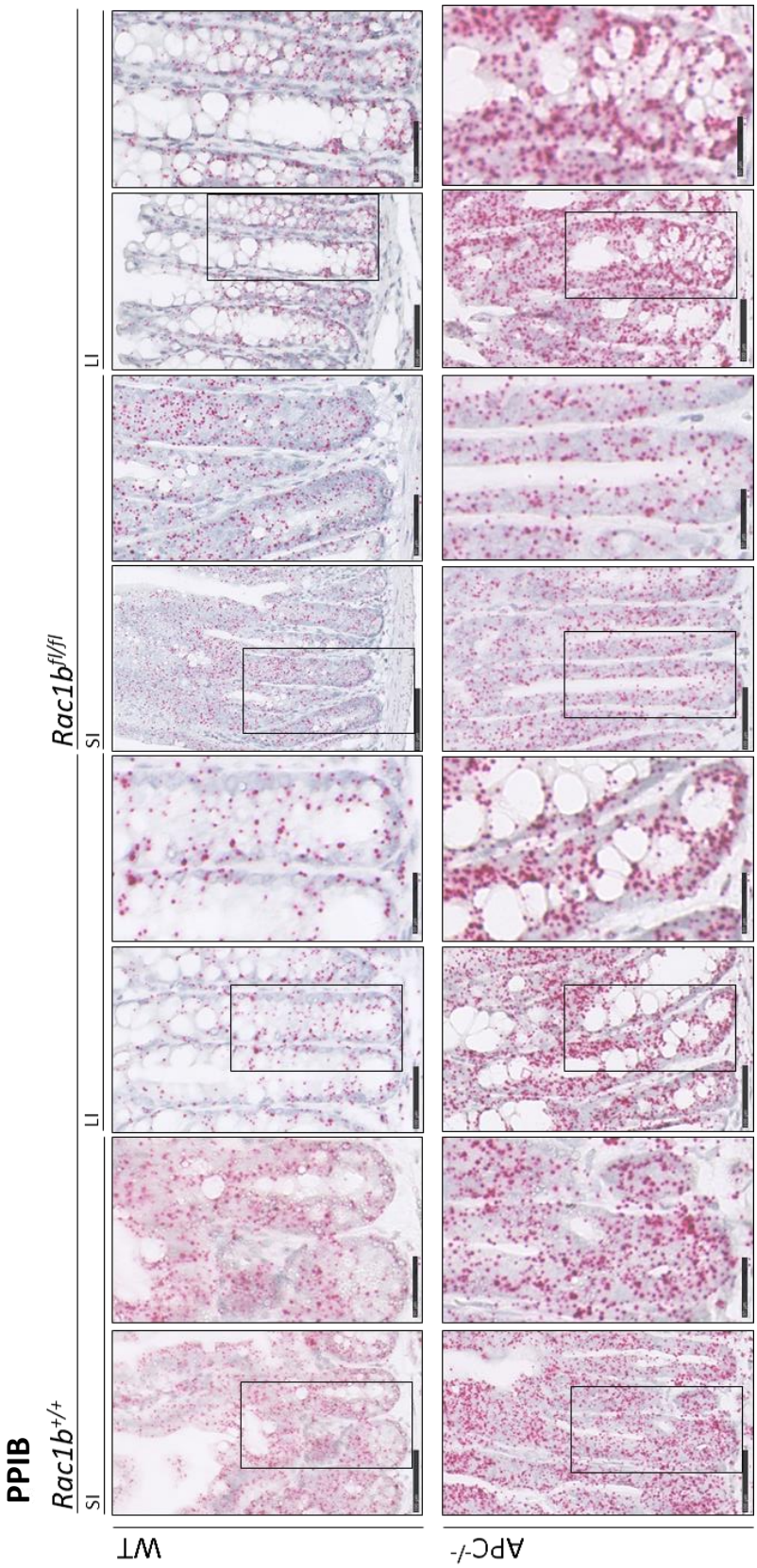
Adenocarcinoma, and Barrett's Epithelium. *Gastroenterology* **141**, 1762–1772 (2011).
492. Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer* **7**, 169–181 (2007).
493. Vaquero, J., Lobe, C. & Fouassier, L. Unveiling resistance mechanisms to EGFR inhibitors in cholangiocarcinoma. *Oncotarget* **9**, 37274–37275 (2018).
494. Nakai, K., Hung, M.-C. & Yamaguchi, H. A perspective on anti-EGFR therapies targeting triple-negative breast cancer. *Am. J. Cancer Res.* **6**, 1609–23 (2016).
495. Gazzaniga, P., Raimondi, C., Urbano, F. & Cortesi, E. EGFR Inhibitor as Second-Line Therapy in a Patient With Mutant *RAS* Metastatic Colorectal Cancer: Circulating Tumor DNA to Personalize Treatment. *JCO Precis. Oncol.* **2**, 1–6 (2018)
496. Chan, D. L. H. *et al.* Epidermal growth factor receptor (EGFR) inhibitors for metastatic colorectal cancer. *Cochrane Database Syst. Rev.* **6** (2017).
497. Rosell, R. *et al.* Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet. Oncol.* **13**, 239–46 (2012).
498. Hu, Z. *et al.* GEP100/Arf6 Is Required for Epidermal Growth Factor-Induced ERK/Rac1 Signaling and Cell Migration in Human Hepatoma HepG2 Cells. *PLoS One* **7**, e38777 (2012).
499. Kurokawa, K., Itoh, R. E., Yoshizaki, H., Nakamura, Y. O. T. & Matsuda, M. Coactivation of Rac1 and Cdc42 at Lamellipodia and Membrane Ruffles Induced by Epidermal Growth Factor. *Mol. Biol. Cell* **15**, 1003–1010 (2004).
500. Halatsch, M.-E. *et al.* Candidate genes for sensitivity and resistance of human glioblastoma multiforme cell lines to erlotinib. *J. Neurosurg.* **111**, 211–218 (2009).
501. Karpel-Massler, G. *et al.* Combined Inhibition of HER1/EGFR and RAC1 Results in a Synergistic Antiproliferative Effect on Established and Primary Cultured Human Glioblastoma Cells. *Mol Cancer Ther.* **9**, 1783-95 (2013).
502. Kaneto, N. *et al.* RAC1 inhibition as a therapeutic target for gefitinib-resistant non-small-cell lung cancer. *Cancer Sci.* **105**, 788–794 (2014).
503. Marcar, L. *et al.* Acquired Resistance of EGFR-Mutated Lung Cancer to Tyrosine Kinase Inhibitor Treatment Promotes PARP Inhibitor Sensitivity. *Cell Rep.* **27**, 3422–3432.e4 (2019).
504. Bos, M. *et al.* PD153035, a tyrosine kinase inhibitor, prevents epidermal growth factor receptor activation and inhibits growth of cancer cells in a receptor number-dependent manner. *Clin. Cancer Res.* **3**, 2099–106 (1997).
505. Esufali, S., Charames, G. S., Pethe, V. V., Buongiorno, P. & Bapat, B. Activation of Tumor-Specific Splice Variant Rac1b by Dishevelled Promotes Canonical Wnt Signaling and Decreased Adhesion of Colorectal Cancer Cells. *Cancer Res.* **67**, 2469–2479 (2007).
506. Henriques, A. F. A., Barros, P., Moyer, M. P., Matos, P. & Jordan, P. Expression of tumor-related Rac1b antagonizes B-Raf-induced senescence in colorectal cells. *Cancer Lett.* **369**, 368–375 (2015).
507. Esufali, S., Charames, G. S. & Bapat, B. Suppression of nuclear Wnt signaling leads to stabilization of Rac1 isoforms. *FEBS Lett.* **16**, 4850-6 (2007).
508. Ahmed, S., Goh, W. I. & Bu, W. I-BAR domains, IRSp53 and filopodium formation. *Semin. Cell Dev. Biol.* **21**, 350–356 (2010).
509. Suetsugu, S., Yamazaki, D., Kurisu, S. & Takenawa, T. Differential roles of WAVE1 and WAVE2 in dorsal and peripheral ruffle formation for fibroblast cell migration. *Dev. Cell* **5**, 595–609 (2003).
510. Kowanetz, K. *et al.* CIN85 Associates with Multiple Effectors Controlling Intracellular Trafficking of Epidermal Growth Factor Receptors. *Mol. Biol. Cell* **15**, 3155–3166 (2004).
511. Wandinger-Ness, A. & Zerial, M. Rab proteins and the compartmentalization of the endosomal system. *Cold Spring Harb. Perspect. Biol.* **6**, a022616 (2014).
512. Snippert, H. J. *et al.* Prominin-1/CD133 Marks Stem Cells and Early Progenitors in Mouse Small Intestine. *Gastroenterology* **136**, 2187–2194.e1 (2009).
513. Hu, T. & Li, C. Convergence between Wnt-β-catenin and EGFR signaling in cancer. *Mol. Cancer* **9**, 236 (2010).
514. Lu, Z. & Hunter, T. Wnt-independent beta-catenin transactivation in tumor development. *Cell*

Cycle **3**, 571–3 (2004).
515. Waterman, H. & Yarden, Y. Molecular mechanisms underlying endocytosis and sorting of ErbB receptor tyrosine kinases. *FEBS Lett.* **490**, 142–152 (2001).
516. Anne J. Ridley. Rho GTPases and actin dynamics in membrane protrusions and vesicle trafficking. *TRENDS Cell Biol.* **16**, (2006).
517. Sorkin, A. & Goh, L. K. Endocytosis and intracellular trafficking of ErbBs. *Exp. Cell Res.* **315**, 683–696 (2009).
518. Huang, F., Goh, L. K. & Sorkin, A. EGF receptor ubiquitination is not necessary for its internalization. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 16904–9 (2007).
519. Sigismund, S. *et al.* Clathrin-independent endocytosis of ubiquitinated cargos. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 2760–5 (2005).
520. Stasyk, T. & Huber, L. A. Spatio-Temporal Parameters of Endosomal Signaling in Cancer: Implications for New Treatment Options. *J. Cell. Biochem.* **117**, 836–843 (2016).
521. Amaddii, M. *et al.* Flotillin-1/reggie-2 protein plays dual role in activation of receptor-tyrosine kinase/mitogen-activated protein kinase signaling. *J. Biol. Chem.* **287**, 7265–78 (2012).
522. Meister, M. *et al.* Regulation of cargo transfer between ESCRT-0 and ESCRT-I complexes by flotillin-1 during endosomal sorting of ubiquitinated cargo. *Nat. Publ. Gr.* **6**, 344 (2017).
523. Jia, D. *et al.* WASH and WAVE actin regulators of the Wiskott-Aldrich syndrome protein (WASP) family are controlled by analogous structurally related complexes. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10442–7 (2010).
524. Steinberg, F. *et al.* A global analysis of SNX27–retromer assembly and cargo specificity reveals a function in glucose and metal ion transport. *Nat. Cell Biol.* **15**, 461–471 (2013).
525. MacDonald, E. *et al.* HRS-WASH axis governs actin-mediated endosomal recycling and cell invasion. *J. Cell Biol.* **217**, 2549–2564 (2018).
526. Suetsugu, S. *et al.* Optimization of WAVE2 complex-induced actin polymerization by membrane-bound IRSp53, PIP(3), and Rac. *J. Cell Biol.* **173**, 571–85 (2006).
527. D'Souza-Schorey, C. & Chavrier, P. ARF proteins: roles in membrane traffic and beyond. *Nat. Rev. Mol. Cell Biol.* **7**, 347–358 (2006).
528. Nie, Z. *et al.* A BAR domain in the N terminus of the Arf GAP ASAP1 affects membrane structure and trafficking of epidermal growth factor receptor. *Curr. Biol.* **16**, 130–9 (2006).
529. Lin, D. *et al.* ASAP1, a Gene at 8q24, Is Associated with Prostate Cancer Metastasis. *Cancer Res.* **68**, 4352–4359 (2008).
530. Hashimoto, S. *et al.* Targeting AMAP1 and cortactin binding bearing an atypical src homology 3/proline interface for prevention of breast cancer invasion and metastasis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7036–41 (2006).
531. Nam, J.-M. *et al.* CIN85, a Cbl-interacting protein, is a component of AMAP1-mediated breast cancer invasion machinery. *EMBO J.* **26**, 647–656 (2007).
532. Müller, T. *et al.* ASAP1 promotes tumor cell motility and invasiveness, stimulates metastasis formation in vivo, and correlates with poor survival in colorectal cancer patients. *Oncogene* **29**, 2393–2403 (2010).
533. Inoue, H., Ha, V. L., Prekeris, R. & Randazzo, P. A. Arf GTPase-activating Protein ASAP1 Interacts with Rab11 Effector FIP3 and Regulates Pericentrosomal Localization of Transferrin Receptor–positive Recycling Endosome. *Mol. Biol. Cell* **19**, 4224–4237 (2008).
534. Hashimoto, S. *et al.* Requirement for Arf6 in breast cancer invasive activities. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6647–52 (2004).
535. Albertinazzi, C., Za, L., Paris, S. & de Curtis, I. ADP-Ribosylation Factor 6 and a Functional PIX/p95-APP1 Complex Are Required for Rac1B-mediated Neurite Outgrowth. *Mol. Biol. Cell* **14**, 1295–1307 (2003).
536. Zhang, B. *et al.* beta-Catenin and ras oncogenes detect most human colorectal cancer. *Clin. Cancer Res.* **9**, 3073–9 (2003).
537. Janssen, K. *et al.* APC and Oncogenic KRAS Are Synergistic in Enhancing Wnt Signaling in Intestinal Tumor Formation and Progression. *Gastroenterology* **131**, 1096–1109 (2006).
538. Pearson, H. B., Phesse, T. J. & Clarke, A. R. K-ras and Wnt signaling synergize to accelerate

prostate tumorigenesis in the mouse. *Cancer Res.* **69**, 94–101 (2009).
539. Tan, X. *et al.* Epidermal growth factor receptor: a novel target of the Wnt/beta-catenin pathway in liver. *Gastroenterology* **129**, 285–302 (2005).
540. Ji, H. *et al.* EGF-induced ERK activation promotes CK2-mediated disassociation of alpha-Catenin from beta-Catenin and transactivation of beta-Catenin. *Mol. Cell* **36**, 547–59 (2009).
541. Lu, Z., Ghosh, S., Wang, Z. & Hunter, T. Downregulation of caveolin-1 function by EGF leads to the loss of E-cadherin, increased transcriptional activity of beta-catenin, and enhanced tumor cell invasion. *Cancer Cell* **4**, 499–515 (2003).
542. Makinoshima, H. *et al.* Epidermal growth factor receptor (EGFR) signaling regulates global metabolic pathways in EGFR-mutated lung adenocarcinoma. *J. Biol. Chem.* **289**, 20813–23 (2014).
543. Bohin, N., Carlson, E. A. & Samuelson, L. C. Stem Cell Reports Report Genome Toxicity and Impaired Stem Cell Function after Conditional Activation of CreER T2 in the Intestine. *Stem Cell Reports*, **11**, 1337-1346 (2018).
544. Bowcutt, R. *et al.* Heterogeneity across the murine small and large intestine. *World J. Gastroenterol.* **20**, 15216–32 (2014).
545. Pećina-Slaus, N. Tumor suppressor gene E-cadherin and its role in normal and malignant cells. *Cancer Cell Int.* **3**, 17 (2003).
546. Huels, D. J. *et al.* E-cadherin can limit the transforming properties of activating b-catenin mutations. *EMBO J.* **34**, 2321–2333 (2015).
547. Means, A. L. *et al.* Epithelial Smad4 Deletion Up-Regulates Inflammation and Promotes Inflammation-Associated Cancer. *Cell Mol Gastroenterol Hepatol.* **24**, 257-276 (2018).
548. Means, A. L. *et al.* Epithelial Smad4 Deletion Up-Regulates Inflammation and Promotes Inflammation-Associated Cancer. *Cell. Mol. Gastroenterol. Hepatol.* **6**, 257–276 (2018).
549. Szigeti, R. *et al.* SMAD4 haploinsufficiency associates with augmented colonic inflammation in select humans and mice. *Ann. Clin. Lab. Sci.* **42**, 401–8 (2012).
550. Ungefroren, H. *et al.* Rac1b negatively regulates TGF-B-induced cell motility in pancreatic ductal epithelial cells by suppressing Smad signalling. *Oncotarget* **5**, 277–290 (2014).
551. Lozano, E., Frasa, M. A. M., Smolarczyk, K., Knaus, U. G. & Braga, V. M. M. PAK is required for the disruption of E-cadherin adhesion by the small GTPase Rac. *J. Cell Sci.* **121**, 933–8 (2008).
552. Freeman, T. J. *et al.* Smad4-Mediated Signaling Inhibits Intestinal Neoplasia by Inhibiting Expression of β-Catenin. *Gastroenterology* **142**, 562–571.e2 (2012).
553. Schwitalla, S. *et al.* Loss of p53 in Enterocytes Generates an Inflammatory Microenvironment Enabling Invasion and Lymph Node Metastasis of Carcinogen-Induced Colorectal Tumors. *Cancer Cell* **23**, 93–106 (2013).
554. Wiener, Z. *et al.* Oncogenic mutations in intestinal adenomas regulate Bim-mediated apoptosis induced by TGF-β. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2229-36 (2014).
555. Markowitz, S. D. & Bertagnolli, M. M. Molecular Basis of Colorectal Cancer. *N. Engl. J. Med.* **361**, 2449–2460 (2009).
556. Ungefroren, H., Witte, D. & Lehnert, H. The role of small GTPases of the Rho/Rac family in TGF-β-induced EMT and cell motility in cancer. *Dev. Dyn.* **247**, 451–461 (2018).
557. Ramesh, S., Wildey, G. M. & Howe, P. H. Transforming growth factor β (TGFβ)-induced apoptosis: The rise and fall of Bim. *Cell Cycle* **8**, 11–17 (2009).
558. Ohgushi, M. *et al.* Transforming Growth Factor β-Dependent Sequential Activation of Smad, Bim, and Caspase-9 Mediates Physiological Apoptosis in Gastric Epithelial Cells. *Mol. Cell. Biol.* **25**, 10017–10028 (2005).
559. Ramjaun, A. R., Tomlinson, S., Eddaoudi, A. & Downward, J. Upregulation of two BH3-only proteins, Bmf and Bim, during TGFβ-induced apoptosis. *Oncogene* **26**, 970–981 (2007).
560. Ramesh, S. *et al.* TGF beta-mediated BIM expression and apoptosis are regulated through SMAD3-dependent expression of the MAPK phosphatase MKP2. *EMBO Rep.* **9**, 990–7 (2008).
561. Wildey, G. M., Patil, S. & Howe, P. H. Smad3 potentiates transforming growth factor beta (TGFbeta)-induced apoptosis and expression of the BH3-only protein Bim in WEHI 231 B lymphocytes. *J. Biol. Chem.* **278**, 18069–77 (2003).
562. Wiener, Z. *et al.* Oncogenic mutations in intestinal adenomas regulate Bim-mediated

apoptosis induced by TGF-β. doi:10.1073/pnas.1406444111
563. Massagué, J. & Gomis, R. R. The logic of TGFβ signaling. *FEBS Lett.* **580**, 2811–2820 (2006).
564. Yu, J. *et al.* Identification of the gene transcription and apoptosis mediated by TGF-β-Smad2/3-Smad4 signaling. *J. Cell. Physiol.* **215**, 422–433 (2008).
565. Ramesh, S., Wildey, G. M. & Howe, P. H. Transforming growth factor β (TGFβ)-induced apoptosis: The rise and fall of Bim. *Cell Cycle* **8**, 11–17 (2009).
566. Reed, K. R. *et al.* A limited role for p53 in modulating the immediate phenotype of Apc loss in the intestine. *BMC Cancer* **8**, 162 (2008).
567. Gonçalves, V., Matos, P. & Jordan, P. The beta-catenin/TCF4 pathway modifies alternative splicing through modulation of SRp20 expression. *RNA* **14**, 2538–49 (2008).
568. Matos, P. *et al.* B-RafV600E Cooperates With Alternative Spliced Rac1b to Sustain Colorectal Cancer Cell Survival. *Gastroenterology* **135**, 899–906 (2008).
569. Ben-Chetrit, N. *et al.* Synaptojanin 2 is a druggable mediator of metastasis and the gene is overexpressed and amplified in breast cancer. *Sci. Signal.* **8**, ra7 (2015).
570. Malecz, N. *et al.* Synaptojanin 2, a novel Rac1 effector that regulates clathrin-mediated endocytosis. *Curr. Biol.* **10**, 1383–1386 (2000).
571. Calvayrac, O. *et al.* The RAS-related GTPase RHOB confers resistance to EGFR-tyrosine kinase inhibitors in non-small-cell lung cancer via an AKT-dependent mechanism. *EMBO Mol Med.* **9**, 238-250 (2017).
572. MacDonald, E. *et al.* HRS-WASH axis governs actin-mediated endosomal recycling and cell invasion. *J. Cell Biol* **217**, 2549–2564 (2018).
573. Moon, B.-S. *et al.* Role of Oncogenic K-Ras in Cancer Stem Cell Activation by Aberrant Wnt/β-Catenin Signaling. *JNCI J. Natl. Cancer Inst.* **106**, (2014).
574. Georgopoulos, N. T., Kirkwood, L. A. & Southgate, J. A novel bidirectional positive-feedback loop between Wnt-β-catenin and EGFR-ERK plays a role in context-specific modulation of epithelial tissue regeneration. *J. Cell Sci.* **127**, 2967–82 (2014).
575. Wellenstein, M. D. *et al.* Loss of p53 triggers WNT-dependent systemic inflammation to drive breast cancer metastasis. *Nature* **572**, 538–542 (2019).
576. Itoh, F. *et al.* The FYVE domain in Smad anchor for receptor activation (SARA) is sufficient for localization of SARA in early endosomes and regulates TGF-beta/Smad signalling. *Genes to Cells* **7**, 321–331 (2002).
577. Chen, Y.-G. Endocytic regulation of TGF-β signaling. *Cell Res.* **19**, 58–70 (2009).
578. Zuo, W. & Chen, Y.-G. Specific activation of mitogen-activated protein kinase by transforming growth factor-beta receptors in lipid rafts is required for epithelial cell plasticity. *Mol. Biol. Cell* **20**, 1020–9 (2009).
579. Dienstmann, R., Salazar, R. & Tabernero, J. Overcoming Resistance to Anti-EGFR Therapy in Colorectal Cancer. *Am. Soc. Clin. Oncol. Educ. B.* **35**, e149–e156 (2015).
580. Woolston, A. *et al.* Genomic and Transcriptomic Determinants of Therapy Resistance and Immune Landscape Evolution during Anti-EGFR Treatment in Colorectal Cancer. *Cancer Cell.* **36**, 35-50 (2019).
581. Lu, Y. *et al.* lncRNA MIR100HG-derived miR-100 and miR-125b mediate cetuximab resistance via Wnt/β-catenin signaling. *Nat. Med.* **23**, 1331–1341 (2017).
582. Zhu, G. *et al.* An EGFR/PI3K/AKT axis promotes accumulation of the Rac1-GEF Tiam1 that is critical in EGFR-driven tumorigenesis. *Oncogene* **34**, 5971–5982 (2015).
583. Wang, F. *et al.* SPSB1-mediated HnRNP A1 ubiquitylation regulates alternative splicing and cell migration in EGF signaling. *Nat. Publ. Gr.* **27**, 540–558 (2017).
584. Faber, A. C. *et al.* BIM Expression in Treatment-Naive Cancers Predicts Responsiveness to Kinase Inhibitors. *Cancer Discov.* **1**, 352–365 (2011).
585. Wiener, Z. *et al.* Oncogenic mutations in intestinal adenomas regulate Bim-mediated apoptosis induced by TGF-β. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2229-36 (2014).

Appendix 1: <u>BaseScope control</u>



<u>Figure appendix 1.1:</u> Positive control for BaseScope staining. SI and LI sections from WT and APC$^{-/-}$ mice bearing double deletion of *Rac1b* (*Rac1b*$^{fl/fl}$) or wild type expression, stained with the PPIB probe. Images are shown at 20X and a selected region is zoom in at 40X. BaseScope staining conducted in collaboration with Colin Nixon.

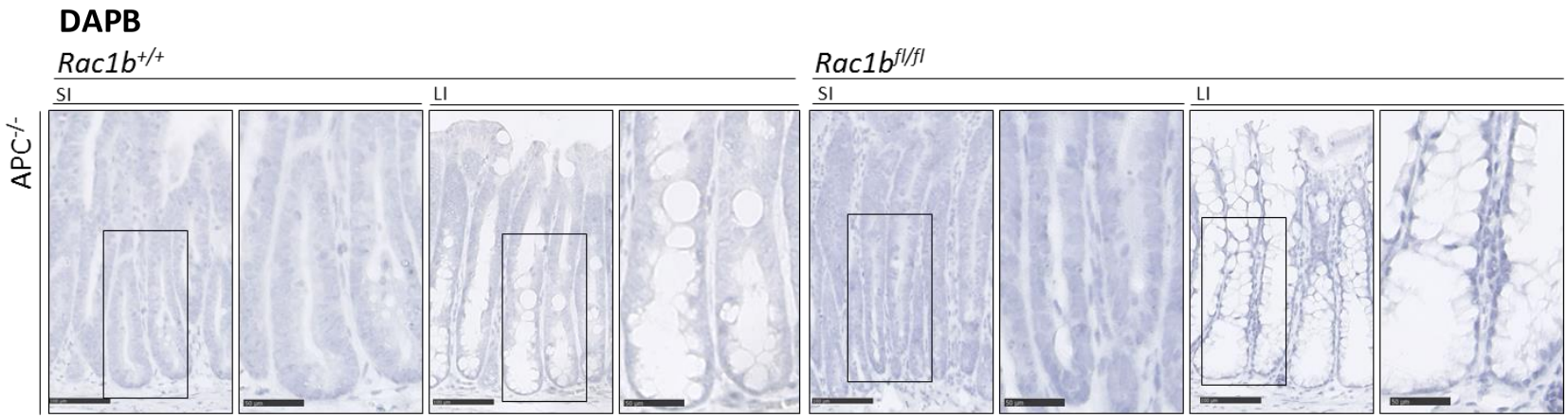


Figure appendix 1.2: Negative control for BaseScope staining. SI and LI sections $APC^{-/-}$ mice bearing double deletion of *Rac1b* (*Rac1b*$^{fl/fl}$) or wild type expression, stained with the DAPB probe. Images are shown at 20X and a selected region is zoom in at 40X. BaseScope staining conducted in collaboration with Colin Nixon.

Appendix 2: RNAseq results

Appendix 2.1: Results RNAseq from section 4.2.3.1

Upregulated genes

gene_name	baseMean	log2FC	pvalue	padj	gene_name	baseMean	log2FC	pvalue	padj	gene_name	baseMean	log2FC	pvalue	padj
Gm45315	317,64	8,74	6,9E-40	1,2E-35	Pira2	51,65	2,22	1,5E-03	4,8E-02	Tmem238l	1464,84	1,02	4,9E-07	2,3E-04
Il22ra2	142,43	7,13	3,6E-05	5,4E-03	Gm4841	13,93	2,21	8,0E-04	3,4E-02	Ldb3	884,97	1,01	1,5E-06	5,8E-04
Fcrl5	60,65	7,03	9,4E-04	3,7E-02	Trim30a	589,37	2,20	2,1E-04	1,5E-02	H2-Q1	903,36	0,99	1,8E-04	1,4E-02
Fgf15	33,65	6,28	1,0E-05	2,4E-03	Poln	11,61	2,20	7,3E-04	3,2E-02	Pdzd3	347,15	0,98	1,6E-04	1,2E-02
Pla2g4c	4269,87	6,07	5,6E-08	5,2E-05	Aqp7	142,57	2,19	4,6E-04	2,5E-02	Apol7a	1114,05	0,97	1,1E-03	4,0E-02
Cacna1i	60,56	6,07	4,5E-04	2,4E-02	Kcnab2	166,10	2,19	4,4E-04	2,4E-02	Rasa4	492,50	0,97	1,5E-03	4,8E-02
Tdg-ps	27,14	5,82	5,2E-05	6,7E-03	Ptpn22	339,70	2,11	8,7E-04	3,5E-02	Espn	1183,81	0,95	4,7E-04	2,5E-02
Ighg3	678,67	5,63	6,8E-05	7,7E-03	Gm48544	17,46	2,05	1,3E-03	4,4E-02	Slc16a5	934,64	0,94	7,2E-04	3,2E-02
Ighm	5122,30	5,56	2,1E-05	4,0E-03	B430010I23Rik	33,36	2,04	1,9E-04	1,4E-02	Naip6	1757,08	0,93	1,2E-03	4,1E-02
Ccr6	182,40	5,48	1,5E-04	1,2E-02	Igkv13-85	606,83	1,99	1,6E-03	5,0E-02	Dtx3l	1848,61	0,92	2,1E-04	1,5E-02
Plb1	537,73	5,33	4,8E-04	2,5E-02	Stk10	460,42	1,99	1,4E-03	4,5E-02	Slc17a1	44,28	0,92	1,1E-03	4,0E-02
Srpk3	64,58	5,28	9,1E-06	2,3E-03	Isg15	501,17	1,98	1,4E-03	4,5E-02	Tstd3	397,15	0,89	1,3E-03	4,3E-02
Kcnj13	13,15	5,27	8,6E-05	8,4E-03	Nos2	233,17	1,97	1,0E-04	9,2E-03	Kbtbd11	1013,87	0,89	6,6E-04	3,0E-02
Ighd	639,65	5,24	7,8E-04	3,4E-02	Gbp8	118,35	1,97	1,6E-03	4,9E-02	Btnl6	2672,53	0,88	1,1E-03	4,0E-02
Smc1b	13,15	5,23	7,1E-05	7,8E-03	Igkv8-21	819,36	1,96	3,4E-04	2,0E-02	Ppfia3	114,68	0,88	1,1E-03	3,9E-02
Il21	12,39	5,00	7,6E-04	3,3E-02	Sp100	1178,57	1,96	4,5E-04	2,4E-02	Mfsd4b3	200,30	0,87	1,6E-03	5,0E-02
Cd22	596,74	4,79	3,8E-04	2,2E-02	Oas1g	447,09	1,93	4,3E-05	6,1E-03	Cldn15	4038,49	0,87	1,3E-05	2,9E-03
Bank1	286,26	4,67	2,5E-04	1,7E-02	Duox2	8370,03	1,91	1,3E-05	2,9E-03	AW112010	1166,51	0,86	1,7E-07	9,7E-05
Map3k19	15,09	4,48	6,9E-04	3,1E-02	Atp8b4	21,07	1,89	7,8E-04	3,4E-02	Trim31	1877,45	0,85	7,4E-05	8,0E-03
Gsdmc3	654,06	4,48	7,9E-04	3,4E-02	Ddx60	896,01	1,85	8,6E-04	3,5E-02	Ifih1	1496,29	0,84	2,1E-04	1,5E-02
Sucnr1	18,84	4,46	1,9E-04	1,4E-02	Klra17	19,98	1,85	6,8E-05	7,7E-03	Evpl	401,58	0,84	1,6E-03	4,9E-02
Ighg2b	12213,24	4,43	1,0E-03	3,9E-02	Oasl2	1804,33	1,78	3,5E-04	2,1E-02	Exoc3l4	690,23	0,83	3,7E-04	2,2E-02
Stab2	30,55	4,16	7,1E-05	7,8E-03	Gm12250	149,62	1,76	1,3E-04	1,1E-02	Gm34923	138,49	0,83	7,1E-05	7,8E-03
Colq	41,73	4,10	7,6E-04	3,3E-02	Mx1	133,20	1,74	4,4E-04	2,4E-02	Dgka	2068,52	0,82	7,5E-04	3,3E-02
Gsdmcl2	179,79	4,06	2,8E-05	4,6E-03	B3gnt5	237,99	1,72	7,1E-04	3,2E-02	Ksr1	1214,60	0,82	9,6E-06	2,4E-03
Gsdmc4	2873,51	4,04	1,6E-03	4,9E-02	Trim34a	452,87	1,72	2,5E-04	1,7E-02	Kcnk5	722,55	0,81	1,0E-03	3,9E-02
Gm19951	68,18	4,04	4,3E-05	6,1E-03	H2-Aa	6358,21	1,70	1,4E-03	4,5E-02	Parp9	1918,59	0,80	4,8E-05	6,4E-03
Il21r	264,73	4,01	7,0E-04	3,1E-02	H2-Eb1	4216,14	1,70	1,6E-03	5,0E-02	Ralgps2	1416,42	0,80	1,0E-03	3,9E-02
4921529L05Rik	17,75	3,99	8,8E-04	3,5E-02	Ighv1-31	119,38	1,68	7,2E-04	3,2E-02	Adap1	1313,82	0,80	3,4E-04	2,0E-02
Gpr174	105,05	3,96	2,6E-04	1,7E-02	Herc6	516,36	1,68	1,6E-03	4,9E-02	Hsph1	1499,95	0,80	1,1E-03	4,1E-02
Gm46392	10,19	3,94	8,1E-05	8,3E-03	Apol10a	3600,63	1,67	1,2E-03	4,1E-02	Slc7a9	2176,23	0,79	1,6E-03	5,0E-02
Saa1	23,55	3,79	5,3E-06	1,7E-03	Il18bp	240,84	1,65	2,7E-04	1,8E-02	Ptprh	2580,02	0,78	2,1E-04	1,5E-02
H3f3aos	32,08	3,63	1,4E-03	4,7E-02	Ighv7-3	441,04	1,63	6,1E-05	7,6E-03	Btnl5-ps	2980,77	0,78	4,8E-04	2,5E-02
Fcrla	377,24	3,62	1,4E-03	4,5E-02	H2-Ab1	5261,56	1,63	1,2E-03	4,3E-02	Anks4b	2937,34	0,77	8,0E-05	8,3E-03
Gm43388	24,53	3,55	8,3E-04	3,4E-02	Irf7	1974,19	1,61	5,1E-05	6,7E-03	Kctd21	241,31	0,77	8,3E-05	8,4E-03
Serpina1b	53,37	3,52	3,0E-05	4,9E-03	Duoxa2	3365,33	1,61	1,1E-04	9,6E-03	Mppe1	322,88	0,77	4,5E-04	2,4E-02
Sell	331,94	3,52	6,4E-05	7,7E-03	Rasgrp4	39,23	1,59	5,9E-04	2,8E-02	Clca3a1	210,90	0,77	1,4E-04	1,1E-02
Ptprcap	105,87	3,51	3,3E-04	2,0E-02	Gm4925	13,74	1,59	1,5E-03	4,8E-02	H2-K1	16271,82	0,76	1,1E-03	3,9E-02
Ptprcap	105,87	3,51	3,3E-04	2,0E-02	Edn3	351,89	1,58	4,8E-04	2,5E-02	Entpd8	1609,57	0,76	1,1E-04	9,7E-03
Tmigd1	333,16	3,44	7,5E-06	2,1E-03	Gm15983	24,32	1,56	5,2E-04	2,6E-02	Clrn3	2864,51	0,76	1,3E-03	4,4E-02
A430078G23Rik	48,71	3,44	5,1E-04	2,6E-02	Ighv9-4	513,03	1,56	3,2E-08	3,6E-05	Mavs	823,26	0,76	8,3E-04	3,4E-02
Tnfsf8	66,65	3,37	6,1E-04	2,8E-02	Asah2	5374,01	1,54	1,5E-03	4,7E-02	Nrarp	183,00	0,76	5,1E-04	2,6E-02
Il27ra	114,86	3,36	2,6E-05	4,5E-03	Tap1	890,30	1,52	9,1E-04	3,6E-02	Cdx1	1295,77	0,75	1,1E-03	3,9E-02
Tbc1d10c	108,59	3,31	5,3E-04	2,6E-02	Apol10b	67,43	1,51	1,1E-04	9,6E-03	Il17rc	903,63	0,74	7,4E-05	8,0E-03
Tbc1d10c	108,59	3,31	5,3E-04	2,6E-02	Gm48111	144,39	1,51	8,1E-04	3,4E-02	Epsti1	713,94	0,74	1,3E-03	4,4E-02
Ikzf1	685,72	3,25	9,1E-05	8,6E-03	Oas1a	982,13	1,48	1,0E-03	3,9E-02	Entpd2	252,33	0,73	1,0E-03	3,9E-02
Napsa	318,36	3,23	1,4E-04	1,1E-02	Ido1	297,71	1,48	5,5E-04	2,7E-02	BC025446	617,02	0,73	5,1E-04	2,6E-02
Gpr18	106,39	3,20	3,5E-04	2,1E-02	Socs1	43,28	1,48	5,5E-04	2,7E-02	Tepsin	464,93	0,73	5,8E-04	2,8E-02
Parvg	248,38	3,18	2,8E-04	1,8E-02	Tnfsf10	523,72	1,47	6,8E-05	7,7E-03	Vil1	22421,24	0,73	3,9E-04	2,2E-02
Gpr183	130,64	3,17	1,4E-04	1,1E-02	Gbp6	270,68	1,46	1,1E-03	4,0E-02	Casp1	1953,37	0,72	9,1E-09	1,5E-05
Myo1g	377,29	3,15	1,6E-04	1,3E-02	Apol7c	431,16	1,46	1,8E-05	3,6E-03	Ms4a8a	2515,18	0,71	1,7E-05	3,5E-03
Oas2	235,65	3,15	9,2E-05	8,7E-03	Zbtb32	44,61	1,43	1,3E-03	4,5E-02	Cfb	2801,40	0,71	5,8E-04	2,8E-02
Car4	616,01	3,12	8,6E-05	8,4E-03	St3gal4	8993,82	1,41	9,8E-04	3,8E-02	Tpcn1	976,05	0,71	4,1E-05	5,9E-03
Ubd	141,04	3,11	1,0E-07	7,6E-05	Duoxa1	82,07	1,40	4,2E-04	2,3E-02	Myrfl	989,55	0,70	3,2E-04	2,0E-02
Lyl1	80,97	3,11	2,6E-04	1,7E-02	Ptger1	42,24	1,40	1,6E-03	4,9E-02	Chic1	301,66	0,67	1,5E-03	4,8E-02
Pglyrp2	52,12	3,07	1,5E-03	4,8E-02	Igtp	616,32	1,40	1,7E-04	1,3E-02	Vipr1	3031,72	0,66	4,4E-05	6,1E-03
Ciita	687,05	3,07	3,2E-05	5,0E-03	Olfr1152	18,62	1,39	9,8E-04	3,8E-02	Camta2	968,89	0,66	3,0E-04	1,9E-02
Gm21596	71,71	3,05	1,4E-21	1,2E-17	Clec4a1	66,98	1,39	1,9E-04	1,4E-02	Myo1a	13689,23	0,66	1,2E-03	4,3E-02
Ccl21a	833,16	3,04	1,2E-04	1,0E-02	2310001H17Rik	266,77	1,39	1,3E-03	4,3E-02	Stambpl1	230,02	0,66	1,0E-03	3,9E-02
S1pr4	54,09	2,99	9,0E-04	3,6E-02	Tmc3	28,00	1,38	1,4E-03	4,5E-02	H2-D1	13723,53	0,66	1,9E-04	1,4E-02
Coro1a	1687,23	2,91	4,9E-04	2,6E-02	Stat1	2139,23	1,36	1,1E-03	4,0E-02	Hepacam2	930,92	0,65	6,7E-05	7,7E-03
Ang4	5443,85	2,90	1,7E-04	1,3E-02	Slc1a1	418,56	1,35	3,0E-04	1,9E-02	Tm6sf2	2656,25	0,64	1,4E-03	4,5E-02
Gm11427	56,11	2,90	9,3E-05	8,7E-03	C2	216,62	1,32	2,0E-05	3,9E-03	Taf4	588,70	0,63	6,6E-05	7,7E-03
Cd69	81,38	2,89	4,0E-04	2,3E-02	Mro	57,75	1,31	1,3E-03	4,5E-02	Zzef1	2999,33	0,62	8,1E-04	3,4E-02
Ighv1-85	579,12	2,88	7,1E-05	7,8E-03	Slc9a9	483,45	1,31	6,3E-06	1,8E-03	Cdx2	1730,67	0,61	2,5E-04	1,7E-02
Slc5a4a	117,60	2,87	9,2E-04	3,7E-02	Ifit1bl1	2004,84	1,29	1,5E-03	4,8E-02	Baiap2l2	1140,32	0,60	3,2E-05	5,1E-03
Aff3	93,74	2,87	1,2E-03	4,3E-02	Slfn2	1086,15	1,29	2,6E-04	1,7E-02	Fam3b	2341,93	0,59	6,1E-04	2,9E-02
Igkv8-16	50,22	2,87	1,3E-04	1,1E-02	Tlr12	106,78	1,29	1,5E-03	4,8E-02	Cyp4f40	922,04	0,59	7,7E-05	8,2E-03
Gm8995	536,80	2,83	4,0E-04	2,3E-02	Syt15	30,37	1,28	1,5E-03	4,8E-02	Psmb10	1578,30	0,59	3,2E-04	2,0E-02
Arhgap4	336,53	2,82	9,5E-04	3,7E-02	Psmb8	1201,20	1,27	5,5E-04	2,7E-02	Ggt6	335,27	0,58	1,0E-03	3,9E-02
Capn13	530,40	2,79	9,3E-06	2,3E-03	Parp14	2052,19	1,27	9,2E-04	3,7E-02	Naip2	1925,94	0,58	2,1E-04	1,5E-02
Skap1	89,30	2,77	1,6E-03	4,9E-02	Ceacam20	1901,96	1,25	2,0E-05	3,9E-03	Rab27a	622,93	0,57	1,1E-03	4,0E-02
H2-Q5	159,27	2,75	1,7E-05	3,5E-03	Gm17132	92,15	1,24	6,3E-05	7,6E-03	Mmaa	295,31	0,57	5,8E-05	7,3E-03
Hvcn1	678,79	2,73	1,4E-03	4,5E-02	Btnl7-ps	552,55	1,23	1,4E-03	4,5E-02	Dok4	1493,46	0,57	4,9E-05	6,5E-03
Arhgdib	1149,11	2,72	7,3E-04	3,2E-02	Gm11775	40,98	1,22	7,8E-05	8,3E-03	Pnpla6	1091,36	0,56	8,2E-04	3,4E-02
Gm5431	188,48	2,72	1,4E-07	8,9E-05	Irgm1	1065,65	1,20	4,7E-05	6,4E-03	Slc7a7	2457,51	0,55	4,6E-04	2,5E-02
Lat	137,01	2,65	1,3E-03	4,4E-02	Rfk	10863,54	1,17	3,5E-04	2,1E-02	Pde9a	1096,66	0,54	5,5E-05	7,0E-03
Gm46432	19,41	2,61	1,3E-08	2,1E-05	Naip5	1961,32	1,15	1,5E-03	4,8E-02	Tle4	996,55	0,54	4,8E-04	2,5E-02
Trim30b	51,74	2,57	7,2E-04	3,2E-02	Ifi47	981,26	1,15	1,0E-04	9,2E-03	Abr	4379,81	0,52	6,9E-04	3,1E-02
Evi2b	210,87	2,56	7,3E-04	3,2E-02	Bst2	501,79	1,14	7,6E-04	3,3E-02	Tff3	11572,00	0,52	5,0E-04	2,6E-02
H2-Q6	771,89	2,52	3,2E-04	2,0E-02	Nlrp9b	420,16	1,10	9,5E-05	8,8E-03	Pex6	615,56	0,51	3,6E-04	2,1E-02
Igkc3	517,18	2,51	5,6E-05	7,0E-03	4933404O12Rik	125,34	1,10	1,1E-03	4,0E-02	Klhl18	578,70	0,51	1,6E-03	4,9E-02
Dennd4b	259,63	2,51	1,3E-03	4,4E-02	Slc9a5	83,20	1,10	7,1E-04	3,2E-02	Usp12	1834,22	0,50	1,3E-03	4,4E-02
H2-Q7	879,07	2,47	4,1E-04	2,3E-02	Tinag	193,78	1,09	6,8E-05	7,7E-03	Mgat3	1582,10	0,50	2,2E-05	4,1E-03
Pvrig	15,60	2,46	8,1E-05	8,3E-03	Batf2	314,01	1,09	4,6E-06	1,6E-03	Stat6	2829,59	0,49	5,9E-04	2,8E-02
Cyp2u1	48,40	2,43	5,3E-04	2,6E-02	Zbp1	891,07	1,09	5,1E-04	2,6E-02	Gna11	6632,24	0,48	2,0E-04	1,5E-02
Cyp2u1	48,40	2,43	5,3E-04	2,6E-02	Bhlha15	211,25	1,08	1,1E-03	4,0E-02	Msantd2	200,85	0,48	1,2E-03	4,3E-02
Oas3	527,60	2,43	8,2E-05	8,3E-03	Ly6a	1747,42	1,08	1,5E-05	3,2E-03	Ckmt1	17043,51	0,46	7,4E-04	3,2E-02
Cd52	1092,89	2,43	1,6E-03	4,9E-02	Uba7	1245,79	1,07	1,5E-03	4,8E-02	Plcb3	6419,24	0,46	3,7E-04	2,2E-02
Itgal	601,84	2,41	1,4E-03	4,5E-02	Parp12	1594,50	1,07	3,9E-04	2,2E-02	Chordc1	1280,54	0,46	1,4E-03	4,5E-02
Gimap4	874,29	2,40	1,2E-03	4,1E-02	Sectm1a	69,50	1,07	1,4E-03	4,7E-02	Rb1	1223,98	0,45	8,6E-05	8,4E-03
Ptpn7	242,71	2,36	8,2E-04	3,4E-02	H2-T23	4095,77	1,06	7,4E-04	3,2E-02	Cpsf4	447,94	0,43	1,2E-03	4,2E-02
Xpnpep2	940,89	2,35	1,9E-04	1,4E-02	Gbp3	303,31	1,06	8,4E-04	3,4E-02	Ccdc93	863,39	0,42	7,8E-04	3,4E-02
Runx3	171,73	2,34	2,7E-04	1,8E-02	Osm	37,80	1,05	8,8E-04	3,5E-02	Uap1	1650,34	0,41	1,2E-03	4,1E-02
Igkv2-112	186,40	2,33	1,3E-05	2,9E-03	Tnfaip8l3	144,99	1,05	1,3E-07	8,9E-05	Fut8	1994,58	0,39	5,7E-04	2,7E-02
Ifi27l2a	220,49	2,30	2,6E-05	4,5E-03	Nuak2	610,39	1,04	3,4E-04	2,0E-02	Rnpepl1	1733,67	0,39	9,2E-04	3,7E-02
Spn	235,03	2,26	9,6E-04	3,7E-02	Csf2rb	544,49	1,03	5,3E-04	2,6E-02	Clsd3	1137,23	0,38	4,9E-04	2,6E-02
Ifi44	455,07	2,25	5,2E-10	1,5E-06	Trim21	265,90	1,02	1,6E-03	4,9E-02	Tbc1d20	2228,12	0,29	9,7E-04	3,8E-02

Downregulated genes

gene_name	baseMean	log2FC	pvalue	padj	gene_name	baseMean	log2FC	pvalue	padj	gene_name	baseMean	log2FC	pvalue	padj
Zfp386	639,94	-0,33	5,1E-04	2,6E-02	Dync2li1	77,34	-0,96	1,1E-03	4,0E-02	Apcdd1	204,10	-1,52	6,5E-04	3,0E-02
Tgfbr2	3915,20	-0,36	1,2E-04	1,0E-02	Hmgb1	6459,81	-0,96	2,1E-09	5,0E-06	Myl9	1705,77	-1,52	8,2E-04	3,4E-02
Trim44	1386,98	-0,39	3,9E-04	2,2E-02	Soga1	155,31	-0,97	8,6E-06	2,2E-03	Ptpru	99,79	-1,56	3,1E-05	4,9E-03
Maco1	977,71	-0,39	5,3E-04	2,6E-02	Fbxo21	543,29	-0,97	1,5E-04	1,2E-02	Fbxo10	168,58	-1,57	3,7E-06	1,3E-03
Atxn10	965,25	-0,41	2,7E-04	1,8E-02	Bambi	178,17	-0,98	2,1E-04	1,5E-02	Hacd4	1026,13	-1,57	6,2E-04	2,9E-02
Rnmt	841,91	-0,41	1,4E-03	4,5E-02	Rgs12	733,75	-0,98	1,0E-03	3,9E-02	Thsd1	160,79	-1,59	1,1E-04	9,6E-03
Alkbh8	454,13	-0,43	1,2E-04	1,0E-02	Rps6ka2	141,98	-1,00	4,7E-08	5,0E-05	Cttnbp2	476,10	-1,60	1,1E-07	7,8E-05
Lrrc45	703,31	-0,43	1,5E-03	4,7E-02	Grik5	98,20	-1,01	2,8E-04	1,8E-02	Ramp3	173,59	-1,62	5,0E-05	6,5E-03
Glul	2766,16	-0,45	1,2E-03	4,1E-02	Acsl4	1239,92	-1,01	3,3E-05	5,1E-03	Prss12	533,63	-1,62	5,1E-05	6,7E-03
Gtf2i	2436,91	-0,46	3,5E-06	1,3E-03	Mapk8ip1	91,67	-1,02	4,3E-07	2,1E-04	Tbx4	49,27	-1,62	3,4E-04	2,0E-02
Rabgap1	1204,39	-0,47	3,6E-05	5,5E-03	Cbarp	99,51	-1,02	5,4E-06	1,7E-03	Mpp2	80,25	-1,63	7,2E-09	1,4E-05
Hsf2	238,36	-0,47	9,0E-04	3,6E-02	Ttbk2	447,47	-1,03	3,6E-06	1,3E-03	Hey1	262,15	-1,63	6,5E-06	1,8E-03
Rfxap	320,65	-0,49	4,0E-05	5,9E-03	Gplhbp1	74,68	-1,03	8,2E-05	8,3E-03	Rcan2	161,17	-1,64	2,1E-04	1,5E-02
Jmy	663,69	-0,50	8,6E-04	3,5E-02	Zc3h12c	109,76	-1,03	1,0E-07	7,6E-05	Robo1	886,61	-1,64	4,9E-04	2,6E-02
1110008P14Rik	280,85	-0,50	9,3E-04	3,7E-02	Arhgef9	112,11	-1,04	1,6E-04	1,3E-02	Cdh3	33,23	-1,65	3,2E-04	2,0E-02
Zfp777	408,94	-0,51	3,8E-04	2,2E-02	Il17rd	293,95	-1,05	1,0E-05	2,4E-03	Sntb1	427,36	-1,65	2,5E-08	3,3E-05
Rgn	130,29	-0,52	1,2E-03	4,1E-02	Fzd6	183,84	-1,05	8,2E-04	3,4E-02	Sptbn2	191,05	-1,65	1,9E-04	1,4E-02
Uri1	948,20	-0,55	4,4E-05	6,2E-03	Fzd6	183,84	-1,05	8,2E-04	3,4E-02	Bean1	116,26	-1,65	6,3E-04	2,9E-02
Slc27a1	353,59	-0,55	4,2E-04	2,3E-02	Sema4c	368,33	-1,05	8,3E-04	3,4E-02	Chn1	158,72	-1,65	1,0E-06	4,5E-04
Phc1	313,42	-0,55	2,5E-04	1,7E-02	Sgpp1	1894,02	-1,06	7,6E-05	8,1E-03	Nipal4	46,95	-1,68	8,7E-05	8,4E-03
Zfp606	248,58	-0,58	1,3E-04	1,1E-02	3110039I08Rik	83,27	-1,07	2,1E-05	4,0E-03	Bex1	105,18	-1,68	7,2E-04	3,2E-02
Id3	1914,93	-0,58	7,9E-04	3,4E-02	Wdr19	233,81	-1,07	9,9E-04	3,8E-02	Gsdma	195,99	-1,69	1,2E-03	4,2E-02
Rfc3	530,62	-0,58	1,1E-03	4,1E-02	Gm2026	129,48	-1,08	1,1E-04	9,9E-03	Prox1	2902,89	-1,69	6,8E-04	3,1E-02
Zfp266	1371,64	-0,59	4,9E-04	2,6E-02	Bmp7	628,41	-1,09	1,4E-03	4,5E-02	Igfbp1	93,03	-1,72	6,9E-04	3,1E-02
Cpq	775,45	-0,60	1,3E-03	4,5E-02	Rcn1	625,81	-1,09	1,5E-05	3,2E-03	Efcab12	12,77	-1,73	1,1E-03	4,0E-02
Zc3h7b	1205,05	-0,60	2,6E-05	4,5E-03	Gprc5b	351,74	-1,09	1,1E-04	9,9E-03	Wnt6	64,13	-1,74	6,2E-05	7,6E-03
Ccdc125	244,58	-0,60	1,6E-03	4,9E-02	Sorcs2	639,46	-1,10	1,1E-04	9,6E-03	Wasf1	65,16	-1,77	5,0E-06	1,7E-03
Prrc2b	3007,24	-0,62	3,8E-07	2,0E-04	Tbc1d16	394,00	-1,10	6,8E-07	3,0E-04	Olfml2a	261,55	-1,78	2,3E-05	4,2E-03
Hdac6	341,88	-0,62	1,7E-07	9,7E-05	Tnfsf18	40,46	-1,14	4,1E-04	2,3E-02	Serpinb9b	16,28	-1,81	9,5E-04	3,7E-02
A430033K04Rik	124,06	-0,63	1,1E-03	3,9E-02	Gm4890	25,64	-1,14	1,6E-03	5,0E-02	Pak3	65,67	-1,81	6,6E-05	7,7E-03
Tmem173	1456,28	-0,65	7,5E-04	3,3E-02	Gm13203	40,30	-1,14	1,1E-03	3,9E-02	Col9a3	594,52	-1,82	1,1E-03	4,1E-02
Aspscr1	1327,27	-0,65	1,7E-04	1,3E-02	Msrb2	58,17	-1,14	2,6E-05	4,5E-03	Tesc	374,29	-1,84	1,1E-03	3,9E-02
Zfp316	119,08	-0,66	3,8E-04	2,2E-02	Tgif2	384,91	-1,15	8,1E-05	8,3E-03	Ascl4	61,87	-1,84	1,8E-05	3,6E-03
Plekhh2	225,44	-0,66	1,2E-04	1,0E-02	Tshz2	525,99	-1,16	5,9E-04	2,8E-02	Fzd3	139,15	-1,86	1,4E-05	3,1E-03
Tns3	3794,27	-0,67	3,8E-04	2,2E-02	Tdrp	67,42	-1,17	6,7E-05	7,7E-03	Nav3	112,86	-1,87	2,6E-09	5,4E-06
Vgll4	1053,37	-0,67	9,4E-04	3,7E-02	Abcc4	1039,56	-1,17	2,7E-05	4,6E-03	Adgrg2	376,44	-1,89	1,1E-11	5,1E-08
Chrnb1	338,97	-0,68	4,0E-04	2,3E-02	AC131339.2	223,63	-1,17	3,4E-04	2,0E-02	Fzd10	152,98	-1,90	5,5E-04	2,7E-02
Ago1	1888,89	-0,69	1,7E-05	3,5E-03	Abca5	214,63	-1,18	2,8E-08	3,3E-05	Cdkn1c	325,13	-1,95	9,4E-05	8,8E-03
Slc9a3	5873,99	-0,69	5,5E-06	1,7E-03	Smtnl2	58,41	-1,18	5,0E-04	2,6E-02	Tmem59l	14,70	-1,96	3,5E-04	2,1E-02
Mfsd12	526,81	-0,71	2,0E-04	1,5E-02	Tmem9	455,64	-1,19	3,6E-04	2,1E-02	Ccdc136	125,08	-1,96	1,7E-05	3,5E-03
Zfp516	256,31	-0,72	7,6E-05	8,1E-03	Plekhb1	58,12	-1,19	2,3E-04	1,6E-02	A730046J19Rik	46,89	-1,96	8,8E-04	3,5E-02
Nceh1	647,56	-0,72	3,1E-06	1,2E-03	Avil	2531,20	-1,19	6,5E-04	3,0E-02	Hs6st2	88,44	-1,97	2,4E-04	1,7E-02
Pik3r1	2543,14	-0,73	3,3E-04	2,0E-02	Fam20c	141,23	-1,20	2,2E-04	1,6E-02	Pkd2l1	33,88	-2,00	1,2E-03	4,2E-02
Map3k15	186,38	-0,74	4,7E-04	2,5E-02	Pycr1	402,99	-1,21	6,9E-04	3,1E-02	9330182L06Rik	85,20	-2,02	1,8E-05	3,6E-03
Fgd1	74,86	-0,74	7,9E-04	3,4E-02	Znrf3	786,19	-1,21	1,1E-06	4,6E-04	Aox4	43,22	-2,05	4,2E-05	6,1E-03
Trim2	2841,32	-0,74	1,0E-03	3,9E-02	Amotl1	948,28	-1,21	9,3E-05	8,7E-03	Svopl	16,43	-2,05	1,2E-05	2,7E-03
Bmp1	779,29	-0,74	4,5E-07	2,2E-04	Gmpr	59,69	-1,21	4,6E-04	2,5E-02	Cnr1	35,30	-2,06	3,1E-04	2,0E-02
AC119928.2	185,73	-0,75	3,2E-04	2,0E-02	Gm15387	1328,27	-1,21	1,2E-11	5,1E-08	4930452B06Rik	19,58	-2,06	7,5E-04	3,3E-02
Mvb12b	446,37	-0,75	8,6E-05	8,4E-03	Pcyt1b	150,94	-1,21	1,6E-03	4,9E-02	Tex15	86,57	-2,07	4,6E-04	2,5E-02
Lhpp	435,00	-0,76	8,0E-04	3,4E-02	Dkk3	1540,79	-1,22	2,0E-07	1,1E-04	T	381,47	-2,07	1,6E-03	4,9E-02
Extl2	225,08	-0,76	6,2E-05	7,6E-03	Nanos1	304,87	-1,25	1,3E-04	1,1E-02	Wnk3	20,80	-2,14	1,3E-03	4,4E-02
Sinhcaf	1305,21	-0,77	3,2E-05	5,0E-03	Neurl1a	147,84	-1,25	9,1E-06	2,3E-03	Cxcl17	39,74	-2,15	1,0E-03	3,9E-02
Bmf	456,27	-0,77	3,2E-04	2,0E-02	Sox5	136,30	-1,25	1,9E-07	1,1E-04	Fgf3	63,20	-2,15	1,3E-06	5,3E-04
Fam124a	132,42	-0,77	1,2E-03	4,1E-02	Dennd2c	56,15	-1,26	1,3E-03	4,4E-02	Nxpe5	17,04	-2,16	1,7E-04	1,3E-02
Amt	213,90	-0,78	2,6E-04	1,7E-02	Fggy	126,93	-1,26	3,7E-05	5,6E-03	Gbx2	73,69	-2,21	1,3E-04	1,1E-02
Usp22	2897,65	-0,78	5,4E-04	2,6E-02	Abca8b	50,24	-1,27	3,3E-06	1,2E-03	Krt12	75,55	-2,26	8,3E-06	2,2E-03
Pde1c	75,30	-0,78	1,1E-03	3,9E-02	Lepr	680,21	-1,28	4,0E-05	5,9E-03	Asb4	622,41	-2,27	1,2E-04	1,0E-02
Nav2	554,36	-0,78	8,6E-05	8,4E-03	Ttll7	321,43	-1,29	1,2E-06	5,1E-04	Glb1l2	121,73	-2,28	1,8E-04	1,4E-02
Zfp518b	173,61	-0,78	1,4E-03	4,5E-02	Clu	33754,46	-1,29	2,5E-05	4,5E-03	Spock2	1829,33	-2,32	8,5E-08	6,9E-05
Itm2c	2269,25	-0,79	9,9E-04	3,8E-02	A830036E02Rik	47,82	-1,30	1,5E-03	4,7E-02	Casq1	70,21	-2,32	2,7E-04	1,8E-02
N6amt1	326,76	-0,79	5,5E-04	2,7E-02	Bdnf	54,54	-1,31	1,1E-03	4,0E-02	2700046A07Rik	19,79	-2,32	5,3E-04	2,6E-02
March9	109,75	-0,79	2,0E-04	1,5E-02	Rasl11b	81,04	-1,31	5,4E-06	1,7E-03	Pcdhga11	19,14	-2,33	6,0E-04	2,8E-02
Tmem198b	181,15	-0,79	1,0E-04	9,3E-03	Myocd	32,02	-1,32	5,5E-04	2,7E-02	Frmpd1	45,75	-2,34	5,4E-06	1,7E-03
Pdgfd	95,76	-0,81	1,1E-03	4,0E-02	Kif1a	140,42	-1,33	2,4E-04	1,6E-02	Abca13	33,97	-2,36	1,1E-03	4,0E-02
Arl4c	887,78	-0,81	4,6E-04	2,5E-02	Slc7a2	101,64	-1,34	5,2E-08	5,2E-05	Pdgfrl	265,04	-2,36	2,4E-04	1,6E-02
Tmem176b	6176,97	-0,82	2,5E-04	1,7E-02	Lbp	219,42	-1,35	1,0E-03	3,9E-02	Igfbp2	99,39	-2,39	2,1E-04	1,5E-02
Atp11a	831,50	-0,83	1,5E-08	2,2E-05	Gstm7	418,62	-1,37	6,4E-06	1,8E-03	Tmem212	11,02	-2,40	2,0E-04	1,5E-02
Tuba1a	2345,13	-0,83	4,4E-04	2,4E-02	H2afy2	171,56	-1,37	8,3E-04	3,4E-02	Adam28	177,74	-2,47	2,2E-04	1,6E-02
Nova2	46,46	-0,84	1,4E-03	4,5E-02	Trpv4	78,63	-1,40	6,1E-04	2,9E-02	Prtg	22,15	-2,48	8,2E-04	3,4E-02
1700052K11Rik	88,37	-0,84	1,1E-03	4,0E-02	Cdh5	1327,33	-1,41	5,1E-04	2,6E-02	Scml2	230,51	-2,49	7,7E-06	2,1E-03
Afap1	232,86	-0,84	2,6E-04	1,7E-02	Morn4	38,66	-1,41	2,9E-04	1,9E-02	Cyp26a1	26,20	-2,52	1,0E-03	3,8E-02
Dcun1d3	486,94	-0,85	6,8E-07	3,0E-04	Mapk4	189,04	-1,42	2,2E-04	1,5E-02	Tmprss11e	63,77	-2,55	9,4E-04	3,7E-02
Adam8	453,45	-0,86	1,5E-03	4,9E-02	Tspyl5	16,35	-1,43	5,5E-04	2,7E-02	Bex4	19,12	-2,60	5,9E-04	2,8E-02
Ube2e2	248,96	-0,87	1,7E-04	1,3E-02	Mamld1	143,14	-1,43	1,5E-03	4,8E-02	Muc5ac	102,53	-2,61	4,5E-05	6,2E-03
Pcdh7	320,32	-0,87	9,5E-04	3,7E-02	Serpinf2	100,88	-1,44	6,6E-05	7,7E-03	Slc44a5	30,11	-2,73	5,3E-04	2,6E-02
Dapk1	241,67	-0,89	3,0E-05	4,9E-03	Ephx4	108,36	-1,45	8,5E-06	2,2E-03	Otof	66,76	-2,76	2,1E-04	1,5E-02
Dach1	181,74	-0,90	2,5E-04	1,7E-02	Ttc12	73,15	-1,45	2,6E-05	4,5E-03	A2m	551,44	-2,81	8,1E-05	8,3E-03
Bmyc	292,56	-0,90	1,2E-05	2,7E-03	Cldn1	88,18	-1,46	2,9E-05	4,8E-03	Pnliprp1	677,98	-2,85	7,1E-11	2,4E-07
Rnf150	413,11	-0,91	1,2E-04	1,0E-02	Ccdc33	145,75	-1,47	2,5E-04	1,7E-02	Smim10l2a	69,18	-2,88	4,8E-05	6,4E-03
Rpgrip1l	171,76	-0,91	8,9E-05	8,5E-03	4732490B19Rik	197,99	-1,48	5,5E-04	2,7E-02	Gm7357	21,34	-2,90	1,0E-03	3,9E-02
Ppp1r9a	447,83	-0,92	1,2E-03	4,1E-02	Gm11992	53,12	-1,48	1,5E-03	4,8E-02	Prox1os	12,92	-2,93	2,3E-05	4,2E-03
Slc25a23	289,56	-0,92	6,9E-05	7,7E-03	Cadm1	632,03	-1,48	3,0E-04	1,9E-02	Sprr2f	78,40	-3,01	9,3E-05	8,7E-03
Slc25a4	1771,66	-0,93	1,5E-07	9,2E-05	Kif5c	113,55	-1,49	9,3E-04	3,7E-02	Cnksr2	11,41	-3,16	5,6E-06	1,7E-03
Nrbp2	409,97	-0,93	3,3E-05	5,1E-03	Nexmif	128,19	-1,49	5,0E-04	2,6E-02	Pnoc	102,07	-3,22	1,5E-03	4,8E-02
Ephb6	97,31	-0,93	1,2E-03	4,3E-02	Fam222a	219,26	-1,49	2,2E-06	8,6E-04	Cldn6	20,30	-4,00	4,6E-05	6,4E-03
Cplx2	166,98	-0,93	6,1E-08	5,4E-05	Ctsl	8728,34	-1,50	1,5E-03	4,8E-02	Slc16a14	46,26	-4,31	8,2E-08	6,9E-05
Fbxo44	133,74	-0,94	1,1E-05	2,7E-03	Limch1	418,96	-1,51	1,2E-04	1,0E-02	Gm4811	15,52	-5,93	1,7E-07	9,7E-05
Ltbp4	1323,70	-0,95	7,8E-04	3,4E-02	Gm47547	236,60	-1,52	6,0E-05	7,4E-03	Zfp935	10,31	-6,80	8,6E-05	8,4E-03

Table appendix 2.1: **List of significant upregulated and downregulated genes comparing APC and APC RAC1B^{KO} tumours**. BaseMean is the average of all normalised count values. Log2FC (log2 fold change) is the change in gene expression reported on a logarithmic scale to base 2. Significant changes are evaluated by adjusted p values below 0.05.

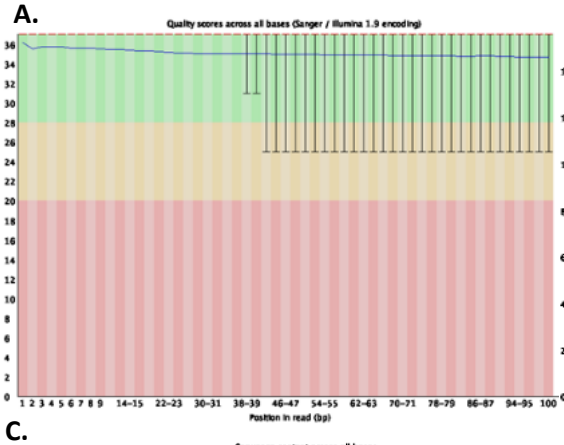
Appendix 2.2: FastQC from RNAseq experiment section 4.2.3.1

Mapping statistics	
APC_1	**APC RAC1B^{KO}_1**
reads processed: 25223430	reads processed: 23816516
reads pseudoaligned: 22803847	reads pseudoaligned: 20139538
percentage pseudoaligned: 90.4	percentage pseudoaligned: 84.6
GC content (%): 48	GC content (%): 48
Sequence length: 100	Sequence length: 100
APC_2	**APC RAC1B^{KO}_2**
reads processed: 24975940	reads processed: 24377962
reads pseudoaligned: 22496979	reads pseudoaligned: 21996774
percentage pseudoaligned: 90.1	percentage pseudoaligned: 90.2
GC content (%): 48	GC content (%): 48
Sequence length: 100	Sequence length: 100
APC_3	**APC RAC1B^{KO}_3**
reads processed: 22127189	reads processed: 24904327
reads pseudoaligned: 19297946	reads pseudoaligned: 21937726
percentage pseudoaligned: 87.2	percentage pseudoaligned: 88.1
GC content (%): 48	GC content (%): 46
Sequence length: 100	Sequence length: 100
APC_4	**APC RAC1B^{KO}_4**
reads processed: 23158769	reads processed: 24853118
reads pseudoaligned: 19959319	reads pseudoaligned: 22238956
percentage pseudoaligned: 86.2	percentage pseudoaligned: 89.5
GC content (%): 48	GC content (%): 48
Sequence length: 100	Sequence length: 100

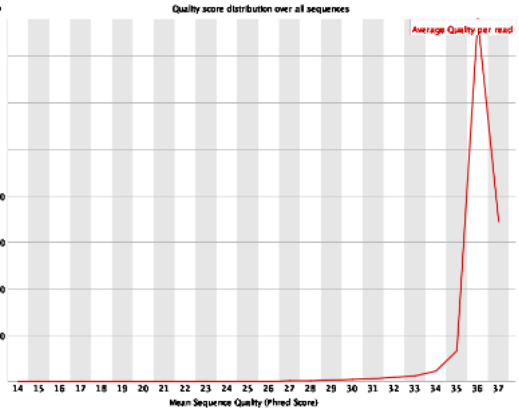
Per base sequence quality (A), per sequence quality scores (B) and per base sequence content (C) modules from the FastQC report are shown below individually per each sample:

APC_1:

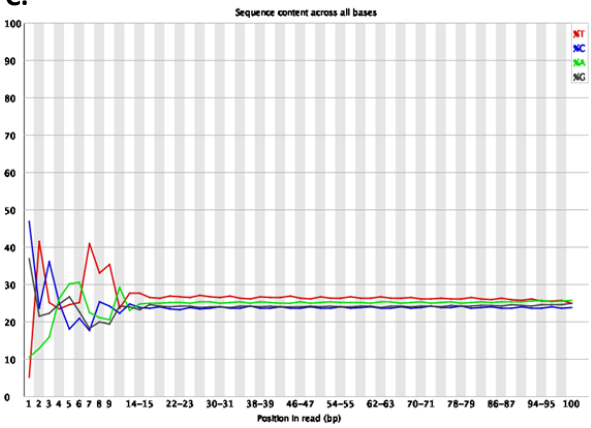
A.



B.

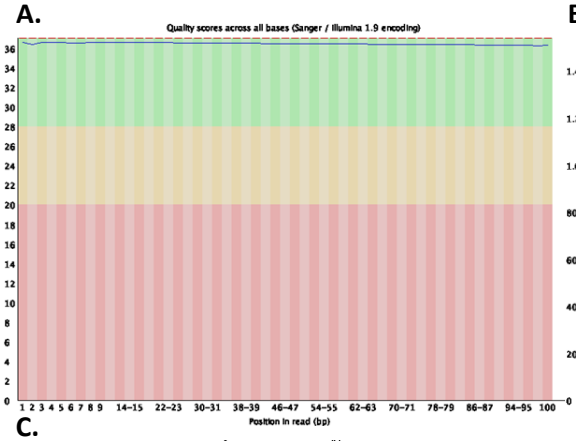


C.

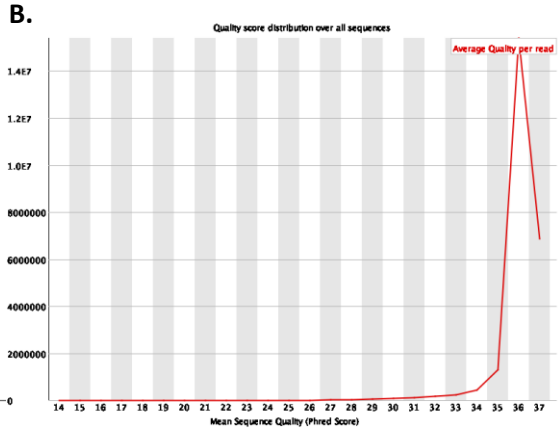


APC_2:

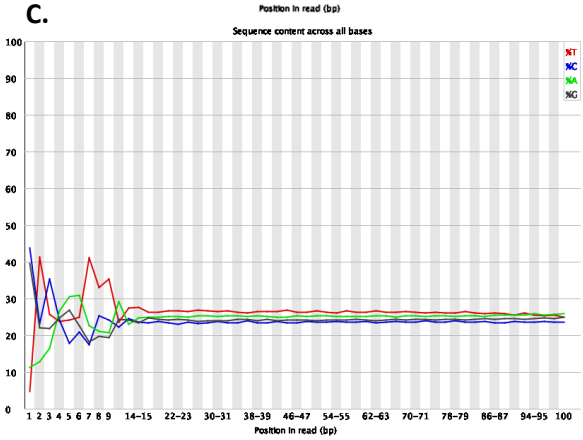
A.



B.

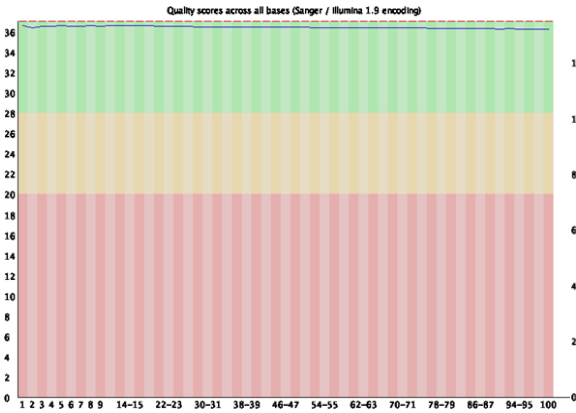


C.

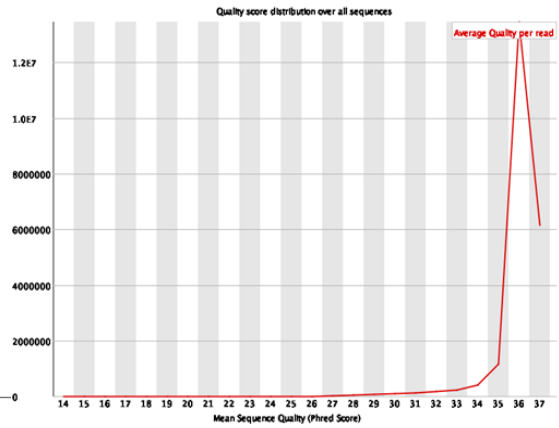


APC_3:

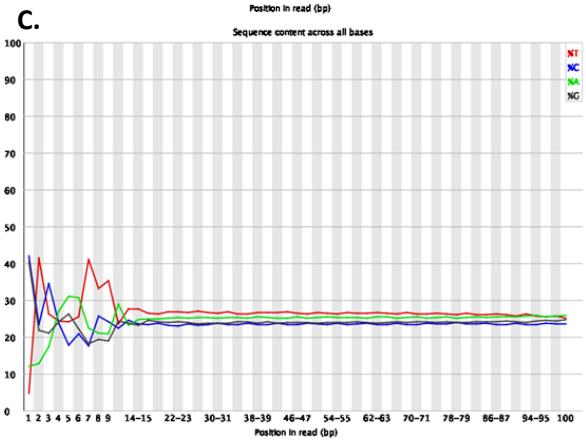
A.



B.

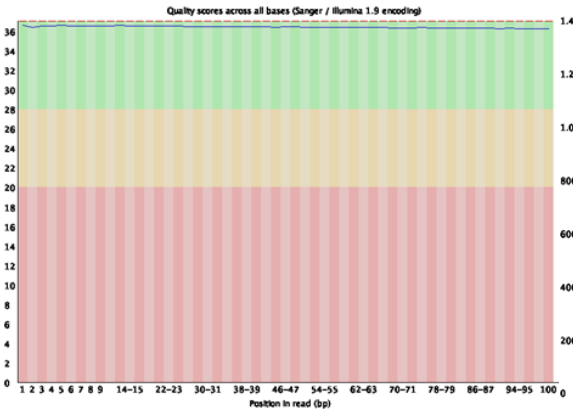


C.

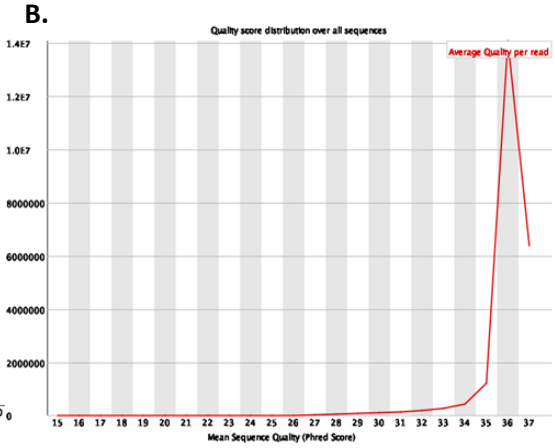


APC_4:

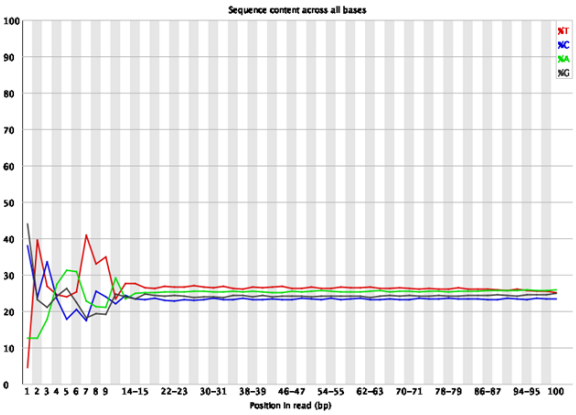
A.



B.

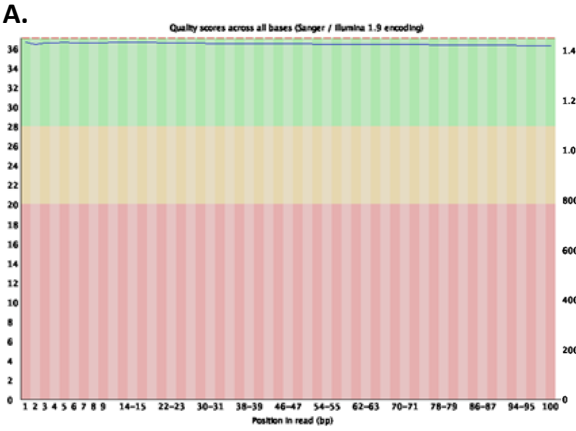


C.

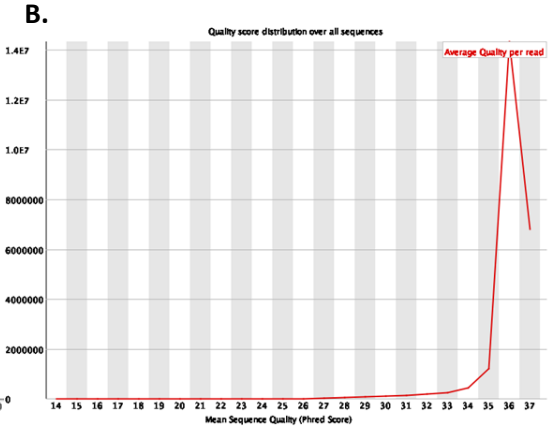


APC RAC1B^{KO}_1:

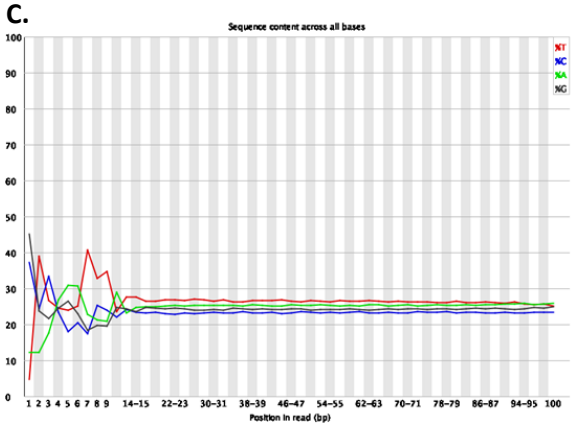
A.



B.

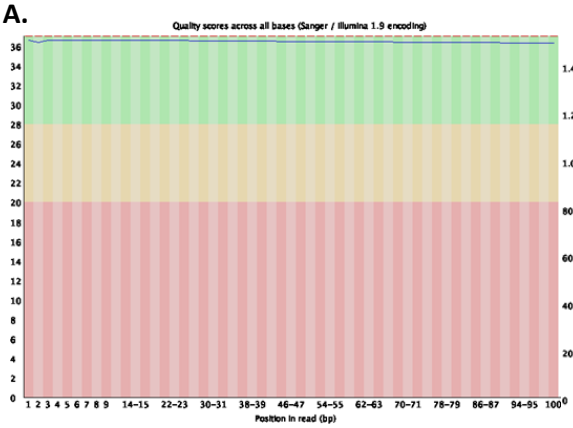


C.

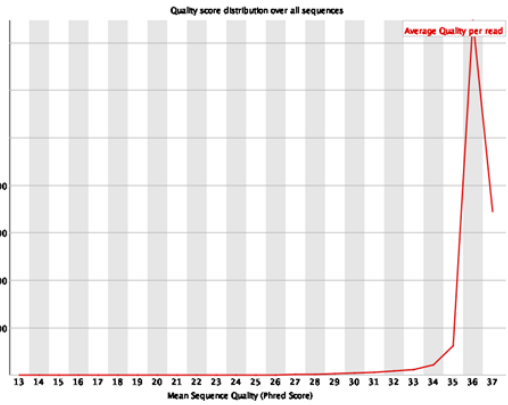


APC RAC1B^{KO}_2:

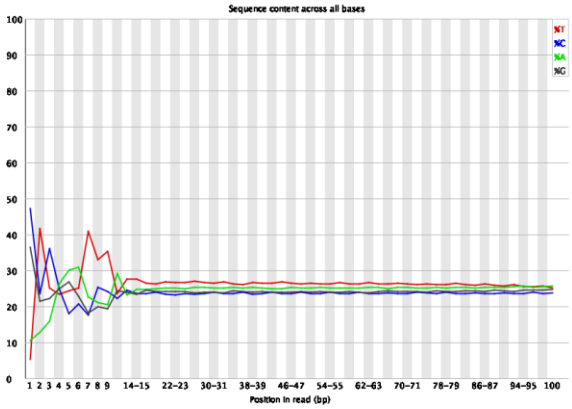
A.



B.

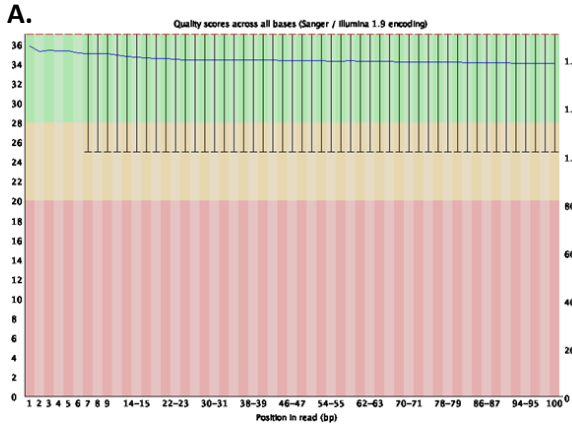


C.

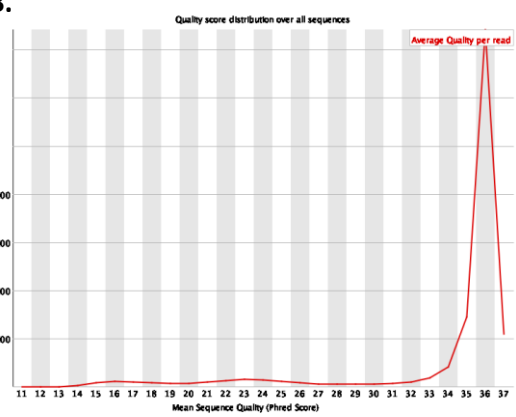


APC RAC1B^{KO}_3:

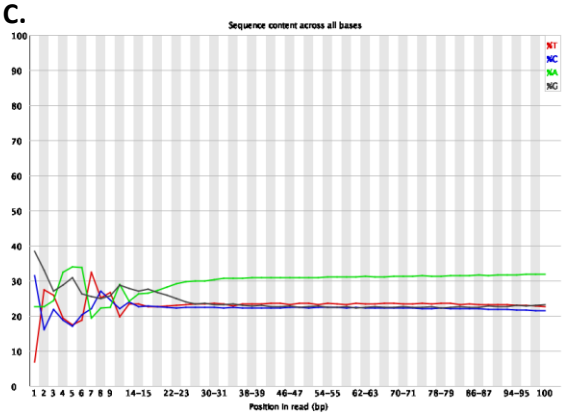
A.



B.

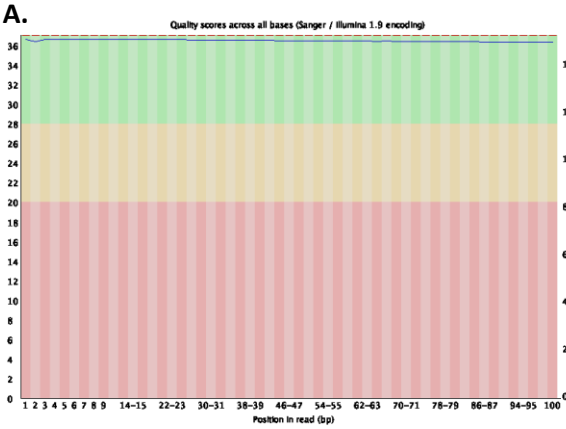


C.

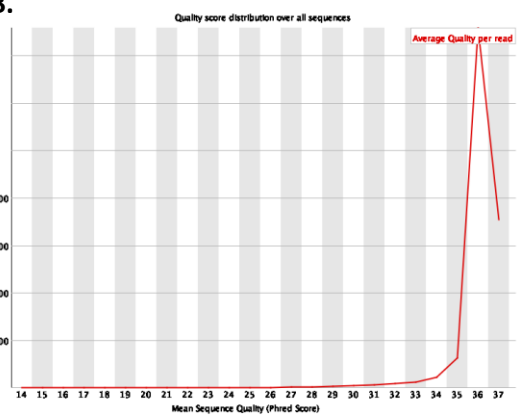


APC RAC1B^{KO}_4:

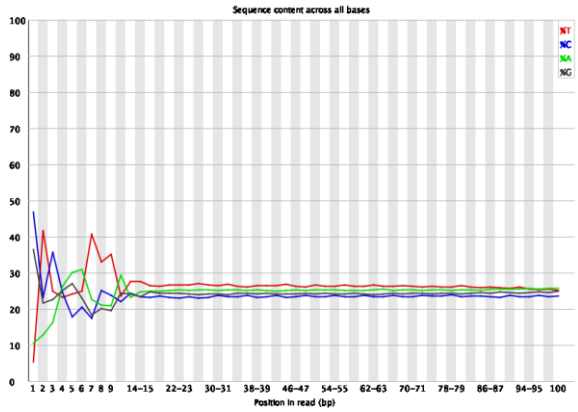
A.



B.



C.



Appendix 2.3: Results RNAseq from section 5.2.1.3

Upregulated genes

gene_name	baseMean	log2FC	pvalue	padj	gene_name	baseMean	log2FC	pvalue	padj
Xist	36,003	9,888	5,0E-05	4,4E-03	Ighv1-83	6,107	1,359	4,5E-04	2,8E-02
Gm16499	3,475	4,854	2,0E-04	1,4E-02	Asb11	7,525	1,357	5,0E-05	4,4E-03
Ighv8-5	127,809	4,611	5,0E-05	4,4E-03	Bex4	1,838	1,355	7,0E-04	4,0E-02
Ighv3-4	14,972	3,706	5,0E-05	4,4E-03	Syt14	0,128	1,353	6,0E-04	3,5E-02
Trpv6	6,429	3,390	5,0E-05	4,4E-03	Igkv9-124	107,654	1,348	5,0E-05	4,4E-03
Neb	3,512	3,364	5,0E-05	4,4E-03	Dsc3	3,702	1,331	1,0E-04	7,6E-03
Igkv4-62	58,895	3,149	5,0E-05	4,4E-03	Fbp1	9,967	1,316	5,0E-05	4,4E-03
Ugt2b38	1,739	2,748	5,0E-05	4,4E-03	Cnr1	0,495	1,300	8,5E-04	4,6E-02
Hoxb9	0,199	2,709	5,0E-05	4,4E-03	Otop1	0,516	1,289	2,0E-04	1,4E-02
Igkv8-34	23,892	2,707	5,0E-05	4,4E-03	Crym	1,455	1,278	1,0E-04	7,6E-03
Shh	0,341	2,683	5,0E-05	4,4E-03	Igkv6-14	74,611	1,273	5,0E-05	4,4E-03
Gm2762	0,221	2,634	1,0E-04	7,6E-03	Dpcr1	12,276	1,271	5,0E-05	4,4E-03
Hoxa9	3,216	2,562	5,0E-05	4,4E-03	Cym	7,494	1,257	5,0E-05	4,4E-03
Ighv7-1	24,958	2,523	5,0E-05	4,4E-03	Mfap3l	0,643	1,230	5,0E-05	4,4E-03
Fmr1nb	1,483	2,501	5,0E-05	4,4E-03	Gm5869	8,526	1,219	1,0E-04	7,6E-03
Spr-ps1	1,458	2,493	5,0E-05	4,4E-03	Ighv1-54	45,181	1,218	5,0E-05	4,4E-03
Ighv1-34	22,936	2,475	5,0E-05	4,4E-03	Ivl	1,110	1,212	3,0E-04	2,0E-02
Cited1	4,949	2,417	5,0E-05	4,4E-03	Oas2	2,796	1,191	5,0E-05	4,4E-03
Alx3	0,497	2,405	5,0E-05	4,4E-03	Gbp3	5,540	1,180	5,0E-05	4,4E-03
Lgr6	0,141	2,289	7,0E-04	4,0E-02	Dmp1	1,637	1,172	5,0E-05	4,4E-03
Cml2	1,484	2,276	5,0E-05	4,4E-03	Nkx6-3	6,819	1,160	5,0E-04	3,0E-02
Ighv1-61	13,521	2,114	5,0E-05	4,4E-03	Ighv5-12	115,519	1,160	5,0E-05	4,4E-03
Vmn1r51	0,227	2,113	6,0E-04	3,5E-02	Ambp	2,216	1,156	3,0E-04	2,0E-02
Ighv1-43	4,154	2,020	5,0E-05	4,4E-03	Paqr9	0,250	1,150	6,0E-04	3,5E-02
Prss56	1,158	1,987	5,0E-05	4,4E-03	Igkv12-98	88,536	1,134	5,0E-05	4,4E-03
Cyp4a10	1,719	1,984	5,0E-05	4,4E-03	Ighv1-52	20,911	1,105	7,0E-04	4,0E-02
Fbxo16	0,228	1,974	8,0E-04	4,4E-02	Slc40a1	332,451	1,098	1,0E-04	7,6E-03
Igkv6-13	71,299	1,969	5,0E-05	4,4E-03	Msx1	17,478	1,094	5,0E-05	4,4E-03
Fam181b	0,151	1,966	3,5E-04	2,2E-02	Iglv2	55,054	1,094	2,0E-04	1,4E-02
Igfbp2	8,076	1,963	5,0E-05	4,4E-03	Cyp2c66	5,313	1,080	1,0E-04	7,6E-03
Ecel1	1,322	1,916	5,0E-05	4,4E-03	Mx1	2,002	1,073	6,0E-04	3,5E-02
Abca13	0,252	1,897	2,0E-04	1,4E-02	H2-Q7	13,846	1,064	1,0E-04	7,6E-03
Ighv1-39	190,111	1,888	5,0E-05	4,4E-03	Rsad2	5,955	1,034	5,0E-05	4,4E-03
Sp8	0,505	1,859	1,0E-04	7,6E-03	Ighv1-81	288,624	1,031	5,0E-05	4,4E-03
Ighv2-5	31,132	1,847	5,0E-05	4,4E-03	Fam189a2	2,174	1,023	6,5E-04	3,8E-02
1700011H14Rik	1,005	1,823	5,0E-05	4,4E-03	Piwil2	1,754	0,998	2,0E-04	1,4E-02
Wnt2	0,146	1,793	2,5E-04	1,7E-02	Agr3	6,863	0,989	6,5E-04	3,8E-02
Igkv4-63	77,340	1,778	5,0E-05	4,4E-03	H2-Q6	6,933	0,974	5,0E-05	4,4E-03
Thegl	0,552	1,759	8,5E-04	4,6E-02	Neto2	4,938	0,970	5,0E-05	4,4E-03
Cyp4f15	0,509	1,753	3,0E-04	2,0E-02	Mamld1	2,358	0,956	2,5E-04	1,7E-02
Hsd17b13	24,358	1,747	5,0E-05	4,4E-03	Adcy1	2,080	0,950	3,5E-04	2,2E-02
Nefl	0,931	1,735	5,0E-05	4,4E-03	Slc23a2	41,006	0,936	8,0E-04	4,4E-02
Neu3	2,912	1,729	5,0E-05	4,4E-03	Prex2	2,168	0,936	5,0E-05	4,4E-03
Ugt2b5	2,235	1,715	5,0E-05	4,4E-03	Cplx2	1,907	0,928	3,5E-04	2,2E-02
Igkv1-133	5,771	1,699	5,0E-05	4,4E-03	Ank	47,613	0,917	5,0E-05	4,4E-03
Igkv4-86	19,659	1,693	5,0E-05	4,4E-03	Zbtb18	27,556	0,917	2,0E-04	1,4E-02
Crb1	0,185	1,656	5,0E-04	3,0E-02	Atp2b1	47,940	0,914	5,0E-05	4,4E-03
Nr0b2	1,586	1,647	5,0E-05	4,4E-03	Rorc	14,291	0,911	5,0E-05	4,4E-03
Cyp2e1	0,312	1,646	5,0E-04	3,0E-02	Cyp2d26	59,648	0,906	5,0E-05	4,4E-03
Gpx6	2,934	1,641	3,5E-04	2,2E-02	Asprv1	6,515	0,878	7,5E-04	4,2E-02
Slc16a9	7,924	1,584	5,0E-05	4,4E-03	Lect2	12,405	0,877	9,0E-04	4,8E-02
Sfrp5	0,998	1,576	5,0E-05	4,4E-03	Dao	7,458	0,868	7,5E-04	4,2E-02
Gm44386	5,983	1,575	5,0E-05	4,4E-03	Robo1	4,278	0,868	5,0E-05	4,4E-03
Muc5ac	0,369	1,565	1,0E-04	7,6E-03	Slc39a14	56,125	0,850	5,0E-04	3,0E-02
Slc37a2	6,292	1,562	5,0E-05	4,4E-03	Scarb1	44,506	0,848	1,0E-04	7,6E-03
Upk3a	1,102	1,542	1,0E-04	7,6E-03	Slc38a1	28,050	0,841	5,0E-05	4,4E-03
Akp3	1324,236	1,520	1,0E-04	7,6E-03	Nadk2	38,138	0,840	1,0E-04	7,6E-03
Cyp1a1	2,476	1,513	5,0E-05	4,4E-03	Cry1	8,775	0,834	1,0E-04	7,6E-03
Ighv1-59	21,197	1,506	5,0E-05	4,4E-03	Myl9	92,478	0,828	3,5E-04	2,2E-02
Igfals	0,631	1,500	5,0E-05	4,4E-03	Apoc3	401,337	0,812	5,0E-05	4,4E-03
Esrrg	0,887	1,492	6,5E-04	3,8E-02	Tmprss15	20,208	0,802	5,0E-04	3,0E-02
Slc30a10	21,669	1,478	5,0E-05	4,4E-03	Kdm6a	8,432	0,799	1,0E-04	7,6E-03
Syn2	1,387	1,463	1,0E-04	7,6E-03	Hmox1	24,439	0,788	8,0E-04	4,4E-02
Ighv1-36	6,301	1,446	9,0E-04	4,8E-02	Slc46a1	85,202	0,753	7,0E-04	4,0E-02
Cybrd1	349,468	1,437	5,0E-05	4,4E-03	Gstm3	23,123	0,745	4,0E-04	2,5E-02
Tnni1	6,588	1,432	3,0E-04	2,0E-02	Cyp3a11	52,253	0,730	5,0E-05	4,4E-03
Ifit2	5,632	1,431	5,0E-05	4,4E-03	Tubb2a	37,819	0,711	2,5E-04	1,7E-02
Igkv3-4	35,137	1,422	5,0E-05	4,4E-03	Tm4sf4	147,691	0,700	3,5E-04	2,2E-02
Ighv1-82	302,032	1,391	5,0E-05	4,4E-03	Tmem65	15,489	0,691	7,0E-04	4,0E-02
Gpc3	5,518	1,386	5,0E-05	4,4E-03	Fam213a	60,922	0,657	6,0E-04	3,5E-02
Pyy	58,946	1,376	5,0E-05	4,4E-03	Slc4a7	32,162	0,652	7,0E-04	4,0E-02
Ighv3-5	15,687	1,369	1,0E-04	7,6E-03	Eif2s3x	64,741	0,587	7,5E-04	4,2E-02
Sbk3	0,341	1,364	7,5E-04	4,2E-02					

Downregulated genes

gene_name	baseMean	log2FC	pvalue	padj	gene_name	baseMean	log2FC	pvalue	padj	gene_name	baseMean	log2FC	pvalue	padj
Aqp1	35,026	-0,644	9,0E-044	8E-02	Mptx2	72,220	-1,106	5,0E-054	4E-03	Igkv4-69	4,298	-1,728	3,0E-042	0E-02
Ano6	18,692	-0,651	5,0E-043	0E-02	Pde2a	9,023	-1,107	5,0E-054	4E-03	Grem1	6,185	-1,739	5,0E-054	4E-03
Rbp1	13,410	-0,673	7,5E-044	2E-02	Reg3g	2231,640	-1,116	5,0E-054	4E-03	Defa21	40,774	-1,747	5,0E-054	4E-03
Mep1b	167,589	-0,697	5,0E-054	4E-03	Aspa	14,044	-1,126	2,0E-041	4E-02	Gif	0,864	-1,765	5,0E-054	4E-03
Slc17a4	13,437	-0,711	7,0E-044	0E-02	Igkv4-57-1	109,955	-1,126	5,0E-054	4E-03	Gm11331	4,518	-1,771	2,0E-041	4E-02
Mgat4c	45,091	-0,717	1,5E-041	1E-02	Ighv2-9-1	40,610	-1,137	2,5E-041	7E-02	Cxcl3	4,195	-1,799	5,0E-054	4E-03
Pla2g4c	61,573	-0,723	3,0E-042	0E-02	Wfdc2	41,920	-1,146	5,0E-054	4E-03	Ighv2-3	17,047	-1,804	5,0E-054	4E-03
Defa24	1246,035	-0,732	5,0E-043	0E-02	Medag	2,707	-1,149	5,0E-054	4E-03	Anpep	158,708	-1,809	5,0E-054	4E-03
Pmaip1	27,954	-0,740	2,5E-041	7E-02	Fbln2	1,372	-1,155	9,5E-045	0E-02	Kcnj16	0,889	-1,810	5,0E-054	4E-03
Soat2	24,334	-0,750	1,5E-041	1E-02	Csf3r	1,290	-1,165	9,0E-044	8E-02	Tff2	891,359	-1,814	5,0E-054	4E-03
Frat2	10,818	-0,751	1,0E-047	6E-03	Arl4d	2,191	-1,172	1,0E-047	6E-03	Nnmt	0,855	-1,817	5,0E-054	4E-03
St3gal4	215,068	-0,754	1,0E-047	6E-03	Duoxa2	150,786	-1,177	5,0E-054	4E-03	Igkv3-1	13,966	-1,834	5,0E-054	4E-03
Pdzrn3	3,110	-0,756	9,5E-045	0E-02	Unc93a	1,238	-1,190	4,5E-042	8E-02	Gsdmc3	5,464	-1,846	5,0E-054	4E-03
Cd14	27,908	-0,768	5,0E-054	4E-03	Arg1	11,975	-1,197	5,0E-054	4E-03	Plet1	57,517	-1,853	5,5E-054	4E-03
Igkv10-96	704,641	-0,778	5,0E-054	4E-03	BC018473	15,924	-1,200	2,5E-041	7E-02	Blk	0,136	-1,860	4,5E-042	8E-02
Maoa	39,717	-0,781	5,0E-054	4E-03	Prss35	0,736	-1,203	3,0E-042	0E-02	Defa-ps6	14,084	-1,862	5,0E-054	4E-03
Gm15284	362,520	-0,802	5,0E-054	4E-03	Scel	0,392	-1,213	9,5E-045	0E-02	A2m	5,482	-1,881	5,0E-054	4E-03
Ly6d	113,954	-0,803	5,0E-054	4E-03	Hp	14,707	-1,217	5,0E-054	4E-03	Trim38	1,742	-1,931	5,0E-054	4E-03
P2ry2	5,710	-0,803	5,5E-043	3E-02	Gm15293	22,511	-1,221	5,0E-054	4E-03	Gp2	3,037	-1,932	5,0E-054	4E-03
Defa-rs1	44,531	-0,816	6,5E-043	8E-02	B3galt5	10,237	-1,233	5,0E-054	4E-03	Mbl2	6,720	-1,950	5,0E-054	4E-03
Spdef	35,677	-0,822	2,5E-041	7E-02	Gkn2	4,528	-1,238	2,5E-041	7E-02	Gm21013	0,404	-1,956	1,5E-041	1E-02
Creb3l1	30,284	-0,828	5,0E-054	4E-03	Ighv1-5	43,862	-1,245	2,5E-041	7E-02	Vnn3	0,162	-1,957	7,0E-044	0E-02
Asah2	26,210	-0,830	5,0E-054	4E-03	Sh2d6	11,743	-1,246	1,0E-047	6E-03	Gm42653	1,340	-1,968	5,0E-054	4E-03
Agr2	976,288	-0,835	1,0E-047	6E-03	Mep1a	4,206	-1,253	1,0E-047	6E-03	Defa22	46,692	-1,994	5,0E-054	4E-03
Slc15a1	29,501	-0,839	1,0E-047	6E-03	Asz1	1,808	-1,268	8,5E-044	6E-02	Cd209c	5,169	-1,998	5,0E-054	4E-03
Muc6	42,168	-0,853	5,0E-054	4E-03	Igkv4-58	23,644	-1,273	8,5E-044	6E-02	Guca2a	133,702	-1,999	5,0E-054	4E-03
Gm37697	29,579	-0,853	5,0E-054	4E-03	1810046K07Rik	0,933	-1,287	4,5E-042	8E-02	Igkv5-45	12,044	-2,003	5,0E-054	4E-03
Atp2c2	4,797	-0,866	1,5E-041	1E-02	Mgp	5,865	-1,288	5,0E-054	4E-03	Slc5a4a	0,882	-2,018	5,0E-054	4E-03
Reg4	270,736	-0,874	5,0E-054	4E-03	Cxcr2	0,454	-1,291	5,5E-043	3E-02	Gsdmcl2	5,895	-2,021	5,0E-054	4E-03
Khdrbs3	8,967	-0,887	2,0E-041	4E-02	G6pc	7,744	-1,298	5,0E-054	4E-03	Saa3	66,097	-2,022	5,0E-054	4E-03
Slc46a3	9,248	-0,889	2,5E-041	7E-02	Apol10a	40,681	-1,298	5,0E-054	4E-03	Angpt4	0,150	-2,032	5,0E-054	4E-03
Slc51b	35,588	-0,894	5,5E-043	3E-02	Ang4	374,899	-1,300	5,0E-054	4E-03	Gm42851	1,439	-2,060	5,0E-054	4E-03
Tmem86a	15,307	-0,895	5,0E-054	4E-03	2610203C20Rik	0,578	-1,308	1,0E-047	6E-03	Ighv5-16	149,611	-2,124	5,0E-054	4E-03
Vwa1	4,405	-0,897	5,5E-043	3E-02	Mmp3	1,970	-1,330	5,0E-054	4E-03	Ighv1-4	21,669	-2,135	5,0E-054	4E-03
Ighv6-3	142,467	-0,901	1,0E-047	6E-03	Clec4d	1,430	-1,344	3,5E-042	2E-02	Retnlg	6,225	-2,182	5,0E-054	4E-03
Gpx2	542,691	-0,907	5,0E-054	4E-03	Kcne4	0,606	-1,345	9,5E-045	0E-02	Muc1	18,648	-2,202	2,0E-041	4E-02
Spink1	427,012	-0,908	5,0E-054	4E-03	Chrnb4	0,264	-1,365	8,5E-044	6E-02	Ighv13-2	65,702	-2,202	5,0E-054	4E-03
Cndp2	58,268	-0,919	5,0E-054	4E-03	Gna14	3,225	-1,375	5,0E-054	4E-03	Cxcl5	17,713	-2,239	5,0E-054	4E-03
St6galnac1	12,565	-0,930	1,0E-047	6E-03	Retnlb	175,105	-1,392	5,0E-054	4E-03	Gm13988	12,960	-2,288	5,0E-054	4E-03
Ighv3-8	75,265	-0,936	3,5E-042	2E-02	Tat	3,266	-1,405	5,0E-054	4E-03	Igkv4-80	60,288	-2,297	5,0E-054	4E-03
Trp53	10,264	-0,937	1,5E-041	1E-02	Ighv1-75	68,757	-1,413	5,0E-054	4E-03	Slc5a4b	1,292	-2,298	5,0E-054	4E-03
Guca2b	185,178	-0,943	5,0E-054	4E-03	1700016C15Rik	1,923	-1,428	2,0E-041	4E-02	Tff1	209,394	-2,332	5,0E-054	4E-03
Ighg2b	27,746	-0,944	5,0E-054	4E-03	Col9a1	0,217	-1,430	7,5E-044	2E-02	Slc5a8	3,613	-2,341	5,0E-054	4E-03
Lcn2	152,002	-0,950	5,0E-054	4E-03	Vnn1	21,624	-1,430	5,0E-054	4E-03	S100a9	19,841	-2,404	5,0E-054	4E-03
Ighg1	21,435	-0,956	5,0E-054	4E-03	Igkv1-88	28,011	-1,440	5,0E-054	4E-03	Il6	2,951	-2,408	5,0E-054	4E-03
Aoc1	43,608	-0,956	5,0E-054	4E-03	Xpnpep2	5,606	-1,442	1,0E-047	6E-03	S100a8	16,340	-2,503	5,0E-054	4E-03
Klk1	91,013	-0,958	5,0E-054	4E-03	Slpi	60,950	-1,449	5,0E-054	4E-03	Ighv1-23	2,486	-2,527	5,0E-054	4E-03
Ptk6	4,609	-0,959	5,0E-054	4E-03	Slc28a1	6,158	-1,460	5,0E-054	4E-03	Igkv4-71	12,834	-2,572	5,0E-054	4E-03
Ace	60,527	-0,965	1,5E-041	1E-02	Col11a2	2,547	-1,462	4,0E-042	5E-02	Cpxm1	0,665	-2,584	5,0E-054	4E-03
Gkn1	71,017	-0,966	5,0E-054	4E-03	Hand2	0,786	-1,469	5,0E-054	4E-03	Cldn10	0,246	-2,593	1,0E-047	6E-03
Clec2h	321,334	-0,973	5,0E-054	4E-03	Slc13a2	6,516	-1,471	5,0E-054	4E-03	Fgf15	0,156	-2,671	5,5E-043	3E-02
Ceacam10	39,897	-0,980	5,0E-054	4E-03	Igkv13-84	45,516	-1,473	5,0E-054	4E-03	Car4	16,593	-2,688	5,0E-054	4E-03
Dram1	2,335	-0,982	8,0E-044	4E-02	Paqr6	0,930	-1,475	9,5E-045	0E-02	Gpbar1	0,303	-2,702	3,5E-042	2E-02
Ighv1-19	88,512	-0,993	5,0E-054	4E-03	Chit1	7,923	-1,504	5,0E-054	4E-03	Defa20	38,460	-2,705	5,0E-054	4E-03
Igkv5-39	296,504	-0,994	5,0E-043	0E-02	Scgb2b20	5,010	-1,509	5,0E-054	4E-03	Pi15	0,318	-2,709	5,0E-054	4E-03
Cd177	39,567	-1,002	5,0E-054	4E-03	Gm7298	0,890	-1,522	5,0E-054	4E-03	Slc5a12	1,081	-2,732	2,0E-041	4E-02
Gsta3	16,373	-1,010	5,0E-054	4E-03	Ighv1-2	3,081	-1,522	8,5E-044	6E-02	Slc9b1	0,235	-2,784	5,5E-043	3E-02
Capn6	1,974	-1,014	3,5E-042	2E-02	Enpep	22,753	-1,525	5,0E-054	4E-03	Prl2c2	1,100	-2,944	5,0E-054	4E-03
Sepp1	689,873	-1,018	5,0E-054	4E-03	Aqp5	55,028	-1,539	5,0E-054	4E-03	Ighv9-4	78,613	-2,976	5,0E-054	4E-03
Cd36	8,815	-1,021	1,0E-047	6E-03	Reg3a	19,463	-1,548	5,0E-054	4E-03	Cd163l1	0,180	-2,987	5,0E-054	4E-03
Steap4	3,355	-1,027	5,0E-054	4E-03	Ccl2	7,779	-1,557	5,0E-054	4E-03	Nts	2,699	-2,997	5,0E-054	4E-03
Eln	3,386	-1,033	5,0E-054	4E-03	Il1f9	0,847	-1,582	5,0E-054	4E-03	Csf3	0,469	-2,999	5,0E-054	4E-03
St8sia6	4,523	-1,035	5,0E-054	4E-03	Capn13	4,668	-1,595	5,0E-054	4E-03	Hormad2	0,165	-3,046	5,5E-043	3E-02
Spink4	3932,290	-1,036	5,0E-054	4E-03	Fgf23	0,426	-1,605	8,0E-044	4E-02	Igkv9-123	2,162	-3,108	1,5E-041	1E-02
A4gnt	3,472	-1,040	1,0E-047	6E-03	Osm	0,542	-1,610	1,5E-041	1E-02	Ighv1-11	1,501	-3,118	8,5E-044	6E-02
Bambi-ps1	9,051	-1,040	6,5E-043	8E-02	Gm42546	4,048	-1,620	5,0E-054	4E-03	Sftpd	1,961	-3,133	5,0E-054	4E-03
Lum	5,563	-1,050	1,0E-047	6E-03	F13a1	2,444	-1,628	5,0E-054	4E-03	Sfrp4	0,263	-3,204	5,0E-054	4E-03
Gm14850	52,496	-1,051	1,5E-041	1E-02	Gsdmc4	42,681	-1,638	5,0E-054	4E-03	Ighd	0,408	-3,308	5,0E-054	4E-03
Irx5	2,812	-1,057	4,5E-042	8E-02	Gcg	2,893	-1,638	5,0E-054	4E-03	Pck1	20,022	-3,365	5,0E-054	4E-03
Dio2	3,458	-1,063	5,0E-054	4E-03	Col8a1	0,227	-1,640	5,0E-054	4E-03	Igkv4-90	191,750	-3,406	5,0E-054	4E-03
Wtip	19,738	-1,067	5,0E-054	4E-03	Igkv6-32	198,134	-1,662	5,0E-054	4E-03	Mfap5	0,483	-3,586	5,0E-054	4E-03
Duox2	60,104	-1,072	5,0E-054	4E-03	Scara5	0,214	-1,669	5,0E-054	4E-03	Slc34a2	5,392	-3,649	5,0E-054	4E-03
Lgi2	0,619	-1,079	3,0E-042	0E-02	Wnt7b	0,433	-1,669	5,0E-054	4E-03	Igkv4-61	60,848	-3,722	5,0E-054	4E-03
Chil4	32,126	-1,08158	5,0E-054	4E-03	Igkv14-130	12,756	-1,67843	5,0E-054	4E-03	Naaladl1	17,796	-3,88029	5,0E-054	4E-03
Nos2	1,012	-1,08596	2,5E-041	7E-02	Clip4	0,271	-1,68091	5,0E-054	4E-03	Osr2	3,205	-4,01003	5,0E-054	4E-03
Reg3b	3708,490	-1,09176	5,0E-054	4E-03	Ccl7	2,776	-1,68356	5,0E-054	4E-03	Ighv9-2	30,632	-4,04211	5,0E-054	4E-03
Lrg1	137,481	-1,09179	5,0E-054	4E-03	Gm12873	1,785	-1,69769	8,0E-044	4E-02	Enpp7	10,500	-4,05615	5,0E-054	4E-03
Adam28	2,267	-1,09201	1,0E-047	6E-03	Gsdmc2	29,420	-1,70721	5,0E-054	4E-03	Ighv11-2	2,150	-4,71452	4,5E-042	8E-02
Il1b	9,272	-1,09434	1,5E-041	1E-02	Ighv8-8	62,617	-1,7098	5,0E-054	4E-03	Uty	3,706	-5,56404	5,0E-054	4E-03
Serpine2	13,187	-1,10427	5,0E-054	4E-03	Igkv3-7	16,340	-1,71436	5,0E-054	4E-03	Kdm5d	2,615	-6,01983	5,0E-054	4E-03

Table appendix 2.2: **List of significant upregulated and downregulated genes comparing APC P53 and APC P53 RAC1B^{OE} tumours**. BaseMean is the average of all normalised count values. Log2FC (log2 fold change) is the change in gene expression reported on a logarithmic scale to base 2. Significant changes are evaluated by adjusted p values below 0.05.

Appendix 2.4: Intratumour heterogeneity from RNAseq experiment section 5.2.3.1

Upregulated genes

gene_name	AP_1 FPKM	AP_2 FPKM	AP_3 FPKM	APR_1FPKM	APR_2 FPKM	APR_3 FPKM
Rac1	305,966	376,083	319,007	333,910	354,559	323,481
Xist	0,101	0,120	0,007	66,106	79,903	69,779
Gm16499	0,395	0,138	0,165	19,031	0,515	0,608
Ighv8-5	12,164	16,530	1,444	646,594	20,673	69,449
Ighv3-4	0,849	4,143	1,402	64,974	17,381	1,084
Trpv6	1,994	0,030	1,335	NA	11,466	2,177
Neb	0,584	0,399	0,883	7,480	8,622	3,104
Igkv4-62	18,051	9,598	8,144	274,055	38,904	4,618
Ugt2b38	0,172	0,056	1,123	6,672	2,364	0,044
Hoxb9	0,048	0,049	0,062	0,048	0,698	0,292
Igkv8-34	0,429	2,968	15,640	86,840	36,057	1,416
Shh	0,114	0,075	0,087	0,222	1,290	0,258
Gm2762	0,027	0,060	0,097	0,433	0,139	0,569
Hoxa9	2,051	0,724	0,019	2,852	1,189	12,458
Ighv7-1	11,558	2,145	8,491	96,910	16,143	14,501
Fmr1nb	0,340	0,150	0,846	0,022	3,671	3,870
Spr-ps1	0,453	0,475	0,392	3,309	3,734	0,387
Ighv1-34	3,149	10,890	6,937	71,672	28,393	16,579
Cited1	2,969	0,591	1,124	5,507	11,097	8,408
Alx3	0,329	0,009	0,136	1,253	0,595	0,661
Lgr6	0,090	0,042	0,012	0,066	0,460	0,176
Cml2	0,892	0,037	0,596	3,997	2,948	0,435
Ighv1-61	8,320	5,491	1,417	47,451	8,673	9,772
Vmn1r51	0,218	0,012	0,026	0,535	0,404	0,168
Ighv1-43	1,106	1,532	2,292	0,581	14,540	4,875
Prss56	0,790	0,237	0,374	0,930	2,124	2,496
Cyp4a10	0,452	0,032	1,597	6,112	1,851	0,272
Fbxo16	0,115	0,036	0,127	0,396	0,259	0,438
Igkv6-13	39,382	10,309	37,343	102,803	181,134	56,822
Fam181b	0,084	0,024	0,076	0,228	0,226	0,268
Igfbp2	6,293	1,570	2,029	8,869	8,563	21,135
Ecel1	0,289	0,006	1,367	4,606	1,495	0,171
Abca13	0,285	0,003	0,032	0,676	0,362	0,152
Ighv1-39	107,390	83,079	52,196	494,394	278,599	125,008
Sp8	0,587	0,005	0,063	0,919	0,794	0,662
Ighv2-5	23,209	4,974	12,436	8,864	122,172	15,136
1700011H14Rik	0,410	0,101	0,818	1,039	3,284	0,377
Wnt2	0,076	0,057	0,064	0,111	0,428	0,141
Igkv4-63	42,954	5,008	56,799	82,103	49,369	227,804
Thegl	0,189	0,169	0,397	0,993	1,362	0,200
Cyp4f15	0,243	0,085	0,370	1,129	1,023	0,204
Hsd17b13	12,544	6,858	14,140	64,183	36,766	11,656
Nefl	0,416	0,107	0,766	0,770	1,085	2,439
Neu3	1,497	0,425	2,126	4,745	6,641	2,038
Ugt2b5	1,672	0,514	0,945	7,499	2,122	0,659
Igkv1-133	3,008	2,207	2,938	4,600	18,085	3,787
Igkv4-86	3,854	12,185	11,818	50,334	13,974	25,786
Crb1	0,225	0,018	0,023	0,248	0,146	0,448
Nr0b2	1,050	0,106	1,148	3,220	3,329	0,665
Cyp2e1	0,169	0,096	0,188	0,602	0,173	0,641
Gpx6	3,731	0,164	0,378	3,900	3,131	6,299
Slc16a9	6,726	0,212	4,956	17,540	16,275	1,837
Sfrp5	0,242	1,033	0,229	0,745	3,181	0,558
Gm44386	4,680	1,697	2,643	17,673	6,728	2,477
Muc5ac	0,449	0,009	0,102	0,264	0,394	0,998
Slc37a2	4,259	0,871	4,420	20,328	6,213	1,661
Upk3a	0,669	0,416	0,605	0,864	2,144	1,915
Akp3	672,606	13,974	1367,580	3029,020	2687,980	174,243
Cyp1a1	0,606	0,060	3,188	6,630	4,063	0,308
Ighv1-59	7,319	11,205	14,600	30,204	27,113	36,741
Igfals	0,285	0,301	0,403	0,674	0,942	1,181
Esrrg	0,947	0,178	0,271	0,381	1,038	2,506
Slc30a10	11,671	4,827	17,851	51,234	35,428	8,999
Syn2	1,619	0,265	0,332	2,221	2,344	1,542
Ighv1-36	1,076	3,887	5,184	23,198	1,929	2,530
Cybrd1	144,601	10,525	410,493	642,649	777,382	111,158
Tnni1	3,530	0,970	6,190	18,035	6,981	3,822
Ifit2	2,059	3,027	4,056	6,956	11,331	6,364
Igkv3-4	36,968	14,956	5,358	3,374	126,775	23,388
Ighv1-82	227,834	69,183	203,196	153,455	980,624	177,894
Gpc3	7,496	0,690	0,975	7,039	10,996	5,913
Pyy	24,785	14,371	59,200	152,259	86,046	17,012
Ighv3-5	13,416	4,155	8,698	30,967	32,775	4,110

gene_name	AP_1 FPKM	AP_2 FPKM	AP_3 FPKM	APR_1FPKM	APR_2 FPKM	APR_3 FPKM
Sbk3	0,409	0,041	0,122	0,526	0,434	0,513
Ighv1-83	1,591	3,147	5,542	8,325	8,281	9,757
Asb11	2,829	0,695	9,151	12,520	18,255	1,702
Bex4	1,156	1,924	0,019	2,050	3,349	2,530
Syt14	0,068	0,089	0,059	0,133	0,285	0,134
Igkv9-124	65,258	38,634	78,278	294,083	95,615	74,057
Dsc3	6,101	0,026	0,192	5,460	5,810	4,621
Fbp1	5,870	8,539	2,727	27,165	11,362	4,137
Cnr1	0,658	0,129	0,072	0,698	0,781	0,634
Otop1	0,355	0,173	0,370	0,343	0,931	0,923
Crym	1,227	0,613	0,709	2,691	2,734	0,755
Igkv6-14	26,787	5,914	98,305	57,794	230,893	27,974
Dpcr1	2,461	0,781	18,341	29,158	20,057	2,858
Cym	4,807	3,243	5,218	13,175	9,886	8,638
Mfap3l	0,655	0,244	0,254	0,763	1,362	0,578
Gm5869	7,936	0,644	6,796	12,373	11,776	11,632
Ighv1-54	25,709	21,554	34,222	98,633	77,546	13,420
Ivl	0,652	0,285	1,071	0,828	1,343	2,482
Oas2	1,059	2,187	1,863	3,190	7,209	1,271
Gbp3	3,031	2,523	4,625	6,315	8,934	7,810
Dmp1	1,307	0,905	0,806	1,708	1,697	3,396
Nkx6-3	0,036	0,027	12,583	15,188	12,958	0,122
Ighv5-12	81,163	103,605	29,509	350,830	79,043	48,966
Ambp	1,455	0,362	2,302	4,122	4,245	0,812
Paqr9	0,170	0,196	0,099	0,198	0,271	0,564
Igkv12-98	72,778	40,415	53,062	18,473	101,037	245,454
Ighv1-52	12,899	7,781	19,138	27,885	27,520	30,242
Slc40a1	198,197	24,828	412,226	611,683	640,407	107,363
Msx1	16,794	1,924	14,731	15,263	29,426	26,731
Iglv2	18,438	10,711	76,237	79,852	34,157	110,932
Cyp2c66	4,271	3,100	2,864	11,942	7,729	1,972
Mx1	1,159	1,027	1,686	2,403	4,455	1,285
H2-Q7	5,446	5,522	15,908	10,553	11,767	33,878
Rsad2	2,582	3,581	5,563	5,629	13,369	5,007
Ighv1-81	92,217	121,899	354,919	260,415	463,131	439,160
Fam189a2	2,147	1,279	0,876	3,361	3,775	1,606
Piwil2	1,308	0,411	1,792	3,581	2,334	1,099
Agr3	5,418	5,021	3,356	7,380	9,336	10,667
H2-Q6	3,263	2,711	8,060	5,941	6,533	15,092
Neto2	4,990	2,420	2,602	6,367	7,082	6,167
Mamld1	2,122	1,372	1,317	2,853	3,406	3,075
Adcy1	1,632	1,238	1,387	1,898	2,537	3,789
Slc23a2	29,538	21,598	33,312	70,978	55,239	35,372
Prex2	1,707	1,054	1,704	2,302	3,429	2,811
Cplx2	1,955	0,970	1,017	2,103	3,138	2,260
Ank	40,294	17,567	41,030	79,516	77,744	29,529
Zbtb18	20,498	19,982	16,774	32,232	55,081	20,767
Atp2b1	36,237	22,232	41,263	83,888	74,113	29,910
Rorc	9,430	12,417	7,927	10,821	32,497	12,657
Cyp2d26	49,433	32,855	42,253	108,205	95,075	30,065
Asprv1	3,028	5,334	5,411	4,381	10,419	10,516
Lect2	11,676	2,761	11,804	14,101	15,965	18,123
Dao	5,765	4,555	5,517	14,049	11,856	3,005
Robo1	4,693	1,985	2,407	4,221	4,725	7,635
Slc39a14	45,046	29,995	45,119	87,700	84,601	44,288
Scarb1	26,985	18,100	50,282	75,642	70,272	25,757
Slc38a1	19,859	19,564	20,867	41,685	40,462	25,863
Nadk2	34,733	16,416	30,872	65,311	55,351	26,144
Cry1	6,462	7,036	5,421	7,214	14,951	11,562
Myl9	130,405	33,374	36,197	117,394	109,230	128,271
Apoc3	283,402	375,334	215,152	779,320	456,090	298,721
Tmprss15	25,111	0,099	18,977	27,589	39,415	10,061
Kdm6a	6,381	5,833	6,247	9,984	11,683	10,463
Hmox1	11,435	7,748	34,584	69,100	9,174	14,591
Slc46a1	87,210	33,436	69,763	146,306	127,358	47,140
Gstm3	22,590	14,526	14,725	54,905	20,499	11,496
Cyp3a11	59,648	30,401	27,858	67,892	92,444	35,274
Tubb2a	23,677	18,745	43,623	57,221	41,755	41,897
Tm4sf4	94,503	99,245	143,927	274,801	170,185	103,484
Tmem65	13,599	10,499	11,454	19,910	20,442	17,032
Fam213a	73,875	37,130	30,841	95,627	76,503	51,555
Slc4a7	30,704	11,330	33,001	57,827	45,183	14,928
Eif2s3x	50,692	56,822	47,701	73,071	79,360	80,800

Downregulated genes

gene_name	AP_1 FPKM	AP_2 FPKM	AP_3 FPKM	APR_1FPKM	APR_2 FPKM	APR_3 FPKM
Aqp1	24,747	75,187	28,201	23,020	22,595	36,405
Ano6	20,515	36,739	11,256	14,284	12,735	16,625
Rbp1	13,341	12,094	24,009	9,390	9,887	11,738
Mep1b	183,284	283,384	155,235	123,513	133,060	127,057
Slc17a4	15,038	27,108	7,901	11,405	10,897	8,276
Mgat4c	47,006	71,910	49,285	25,530	44,865	31,949
Pla2g4c	58,140	70,309	101,586	9,350	78,574	51,479
Defa24	1441,080	2572,100	653,536	809,421	926,614	1073,470
Pmaip1	36,204	46,093	22,619	30,192	19,059	13,555
Soat2	26,456	46,806	18,287	19,068	16,584	18,802
Frat2	16,565	14,006	10,148	9,022	6,630	8,539
St3gal4	272,052	296,507	241,530	141,489	199,101	139,725
Pdzrn3	2,847	5,855	3,018	2,089	1,886	2,965
Cd14	27,781	32,230	45,478	18,847	18,576	24,536
Igkv10-96	643,722	470,412	1556,490	550,775	635,178	371,273
Maoa	54,109	74,333	22,182	35,828	27,550	24,298
Gm15284	461,540	791,105	129,593	235,784	314,278	242,819
Ly6d	70,671	172,897	191,022	67,874	84,888	96,374
P2ry2	5,014	10,203	6,562	3,758	3,249	5,476
Defa-rs1	58,233	105,222	6,963	35,819	40,887	20,062
Spdef	23,657	74,788	38,268	21,974	16,338	39,035
Creb3l1	23,264	60,131	32,825	18,809	18,787	27,888
Asah2	27,990	60,008	12,651	20,216	18,737	17,658
Agr2	783,043	1816,220	1153,860	677,999	555,059	871,546
Slc15a1	37,107	53,704	22,727	22,876	22,176	18,420
Muc6	23,166	71,672	68,004	29,027	23,707	37,433
Gm37697	25,847	60,969	27,427	20,732	20,148	22,349
Atp2c2	4,061	9,419	5,105	3,396	3,647	3,156
Reg4	248,125	664,870	138,057	160,143	127,250	285,967
Khdrbs3	7,912	22,964	4,041	6,179	5,621	7,086
Slc46a3	8,757	18,344	8,928	4,876	7,826	6,756
Slc51b	31,419	83,818	23,580	26,232	24,618	23,858
Tmem86a	14,012	37,160	8,552	7,804	8,300	16,013
Vwa1	5,574	8,420	3,203	3,308	2,470	3,458
Ighv6-3	80,169	186,931	289,561	125,210	125,128	47,799
Gpx2	525,291	902,388	696,091	308,104	425,644	398,627
Spink1	448,321	978,167	244,920	266,197	329,298	295,169
Cndp2	63,245	129,902	35,496	37,911	33,938	49,116
St6galnac1	8,813	29,061	11,561	8,686	11,914	5,354
Ighv3-8	218,500	5,887	72,202	38,118	91,587	25,298
Trp53	20,281	7,629	12,550	7,047	6,139	7,940
Guca2b	208,846	486,293	35,716	127,047	87,190	165,974
Ighg2b	15,900	46,006	47,641	19,052	22,061	15,816
Lcn2	119,087	193,239	288,583	73,745	69,860	167,497
Ighg1	9,578	44,513	30,766	6,564	10,660	26,527
Aoc1	53,315	91,063	28,286	32,002	22,862	34,117
Klk1	83,187	155,970	121,375	53,268	67,298	64,981
Ptk6	5,874	8,206	4,181	2,719	3,552	3,125
Ace	87,485	123,220	29,463	24,650	43,275	55,070
Gkn1	94,836	146,610	40,363	26,682	35,504	82,109
Clec2h	475,457	497,216	304,552	241,102	234,878	174,794
Ceacam10	28,425	66,400	64,031	16,857	24,739	38,927
Dram1	1,608	1,788	5,903	1,120	2,038	1,551
Ighv1-19	110,492	43,461	199,537	48,960	74,396	54,228
Igkv5-39	79,917	84,778	1019,750	386,586	101,415	106,583
Cd177	13,431	121,319	23,595	17,114	12,572	49,372
Gsta3	11,843	6,784	47,021	9,958	16,886	5,744
Capn6	1,938	4,858	1,125	1,244	0,809	1,870
Sepp1	745,527	1679,900	345,220	396,472	586,885	385,238
Cd36	11,303	20,788	3,342	5,145	3,283	9,027
Steap4	5,255	3,052	5,197	1,809	1,989	2,829
Eln	3,136	4,275	6,238	2,420	1,384	2,866
St8sia6	4,599	6,446	7,196	2,368	3,087	3,445
Spink4	3784,720	6978,040	5094,920	2109,510	2680,260	2946,280
A4gnt	2,284	6,169	5,561	1,196	2,624	2,997
Bambi-ps1	8,722	18,538	9,274	4,954	4,730	8,086
Lum	5,056	9,446	8,008	2,661	3,387	4,822
Gm14850	59,735	140,425	12,270	29,846	32,696	40,004
Inx5	1,263	9,780	0,354	1,542	2,491	1,444
Dio2	4,787	5,030	4,217	2,197	0,687	3,832
Wtip	21,085	31,658	27,416	10,957	11,060	16,251
Duox2	49,166	102,723	92,504	21,785	54,062	40,382
Lgi2	0,596	0,904	1,021	0,316	0,393	0,484
Chil4	10,286	7,971	112,644	2,084	29,131	30,637
Nos2	0,678	2,454	0,997	0,300	0,622	1,023
Reg3b	4723,310	6243,060	4178,680	984,811	3674,050	2447,010
Lrg1	122,224	190,499	248,739	80,946	49,433	133,047
Adam28	4,302	0,189	4,770	1,487	1,486	1,371
Il1b	12,005	15,315	10,566	5,202	3,582	8,960
Serpine2	17,077	23,739	13,186	7,474	5,116	12,528
Mptx2	111,550	166,812	17,526	41,857	31,188	64,389
Pde2a	7,329	20,406	9,241	5,188	5,231	6,742
Reg3g	1964,700	4141,150	3056,390	856,871	2066,540	1304,180
Aspa	24,345	21,213	12,224	10,063	8,192	8,225
Igkv4-57-1	108,220	28,498	315,730	27,286	67,403	112,593
Ighv2-9-1	100,020	29,168	38,325	27,098	32,152	16,899
Wfdc2	18,872	62,704	91,673	21,357	37,619	19,299
Medag	2,226	5,473	3,495	1,712	1,281	2,056
Fbln2	1,623	2,160	1,897	0,894	0,822	0,834
Csf3r	1,315	1,393	2,646	0,799	0,678	0,910
Arl4d	3,002	5,340	0,761	1,364	1,120	1,557
Duoxa2	122,267	245,297	259,727	51,040	116,588	109,795
Unc93a	1,316	3,490	0,357	1,120	0,578	0,565
Arg1	7,881	26,083	16,066	5,135	7,526	9,157
BC018473	10,122	46,544	9,903	10,175	11,602	7,201
Prss35	0,566	0,618	1,896	0,376	0,696	0,265
Scel	0,244	0,887	0,511	0,181	0,180	0,347
Hp	16,146	17,339	28,222	7,283	3,528	15,724
Gm15293	22,928	67,529	4,064	9,851	13,260	17,436
B3galt5	21,164	7,098	14,830	3,885	10,890	3,554
Gkn2	7,080	6,541	5,460	0,801	1,913	5,377
Ighv1-5	72,379	21,840	90,869	11,123	17,272	49,686
Sh2d6	10,962	17,759	20,846	5,928	6,800	8,165
Mep1a	2,476	13,368	1,935	3,168	2,698	1,594
Asz1	1,623	3,709	2,335	1,340	1,236	0,608
Igkv4-58	51,332	7,940	41,068	6,261	29,943	5,321
1810046K07Rik	1,285	1,977	0,710	0,605	0,732	0,289
Mgp	4,714	12,143	8,107	2,892	3,456	3,876
Cxcr2	0,503	0,557	0,876	0,178	0,144	0,469
G6pc	6,661	24,863	1,506	3,519	4,759	5,159
Apol10a	33,083	109,870	30,576	14,242	29,072	27,243
Ang4	426,356	527,777	645,522	68,904	302,282	278,554
2610203C20Rik	0,293	0,496	1,682	0,269	0,355	0,373
Mmp3	1,868	2,442	4,148	1,475	0,766	1,124
Clec4d	1,712	2,310	2,133	0,571	0,677	1,177
Kcne4	0,735	1,409	0,464	0,304	0,122	0,600
Chrnb4	0,184	0,741	0,215	0,180	0,208	0,054
Gna14	2,218	8,592	3,158	2,027	1,626	1,732
Retnlb	163,299	40,589	556,829	41,096	186,741	62,078
Tat	2,151	9,509	2,566	0,964	1,239	3,169
Ighv1-75	92,555	173,221	34,164	30,732	42,887	38,983
1700016C15Rik	1,408	6,025	0,979	0,923	0,852	1,352
Col9a1	0,073	0,710	0,166	0,069	0,071	0,212
Vnn1	12,401	69,256	12,971	8,379	17,137	9,598
Igkv1-88	13,219	76,942	32,633	6,495	37,028	1,746
Xpnpep2	4,757	16,026	3,807	2,826	3,785	2,436
Slpi	52,643	76,546	138,465	27,657	34,336	36,055
Slc28a1	9,678	13,664	3,758	3,643	2,410	3,797
Col11a2	1,837	6,122	3,252	1,568	0,952	1,549
Hand2	0,596	0,921	1,948	0,325	0,492	0,434
Slc13a2	8,061	19,715	0,958	5,460	1,191	3,714
Igkv13-84	75,779	37,614	87,378	35,544	30,542	6,241
Paqr6	0,383	2,995	0,726	0,667	0,470	0,339
Chit1	5,164	19,247	10,735	3,263	3,698	5,434
Scgb2b20	7,624	9,864	4,755	2,323	2,333	3,158
Gm7298	1,644	0,004	2,313	0,585	0,507	0,287
Ighv1-2	3,971	1,369	8,369	0,462	2,464	1,849
Enpep	33,064	64,547	3,702	8,605	8,622	17,980
Aqp5	41,653	130,409	73,575	25,359	22,109	37,063
Reg3a	20,966	56,245	9,802	7,488	9,932	12,342
Ccl2	11,487	13,705	9,641	3,755	2,300	5,786
Il1f9	0,716	0,757	2,334	0,429	0,437	0,405
Capn13	4,320	11,995	4,729	1,198	2,984	2,785
Fgf23	0,114	1,623	0,188	0,161	0,051	0,421
Osm	0,749	1,122	0,577	0,306	0,175	0,320
Gm42546	1,268	16,147	0,910	1,916	1,966	2,080
F13a1	2,925	7,851	0,304	1,500	0,438	1,647
Gsdmc4	61,994	85,526	46,282	0,735	23,860	37,692
Gcg	3,534	8,550	1,052	1,094	1,261	1,866
Col8a1	0,234	0,637	0,159	0,102	0,104	0,124
Igkv6-32	80,427	169,848	653,031	139,634	56,052	89,811
Scara5	0,346	0,191	0,439	0,081	0,079	0,147
Wnt7b	0,242	0,347	1,389	0,113	0,308	0,201
Igkv14-130	8,076	26,543	23,698	7,476	8,009	2,734
Clip4	0,407	0,766	0,068	0,116	0,126	0,145
Ccl7	3,398	5,287	4,019	1,415	0,709	1,831
Gm12873	0,705	2,242	5,238	0,265	1,529	0,729
Gsdmc2	38,723	56,728	39,686	0,753	15,474	25,159
Ighv8-8	116,824	25,726	145,190	29,210	45,273	13,480
Igkv3-7	25,893	27,770	21,480	3,162	6,927	12,811
Gm4251	0,287	6,338	0,337	0,558	0,487	0,625
Ighv5-16	390,877	198,170	141,145	101,383	41,716	24,376
Ighv1-4	69,378	14,294	22,240	11,147	6,453	6,505
Retnlg	7,879	14,818	7,906	2,196	1,392	3,157
Muc1	0,867	0,995	90,048	1,971	16,950	1,058
Ighv13-2	97,048	20,618	206,165	50,773	11,392	8,215
Cxcl5	16,178	15,381	56,144	4,880	5,916	7,778
Gm13988	2,921	59,600	2,020	6,512	2,764	3,942
Igkv4-80	128,654	43,598	128,322	28,662	29,371	3,122
Slc5a4b	0,438	5,746	0,257	0,436	0,346	0,528
Tff1	50,841	821,015	176,387	48,728	77,985	81,408
Slc5a8	1,204	14,769	2,134	0,218	0,892	2,464
S100a9	22,149	46,274	31,704	7,289	2,172	9,460
Il6	4,128	9,015	1,759	0,487	0,377	1,943
S100a8	19,223	37,727	26,393	5,739	1,764	7,195
Ighv1-23	4,351	7,869	0,489	0,750	1,243	0,212
Igkv4-71	4,785	49,503	11,630	6,882	3,564	0,640
Cpxm1	0,182	1,546	1,692	0,250	0,097	0,223
Cldn10	0,087	1,118	0,061	0,066	0,105	0,038
Fgf15	0,054	0,706	0,047	0,018	0,064	0,045
Car4	17,185	64,711	4,283	1,705	3,768	7,904
Gpbar1	0,136	0,494	0,947	0,082	0,097	0,064
Defa20	10,565	187,905	1,607	4,823	13,581	12,278
Pi15	0,275	1,157	0,221	0,086	0,069	0,097
Slc5a12	0,244	5,382	0,011	0,022	0,216	0,610
Slc9b1	0,471	0,323	0,436	0,056	0,079	0,045
Prl2c2	0,213	3,146	2,485	0,233	0,378	0,148
Ighv9-4	391,188	15,438	11,846	27,271	15,713	10,219
Cd163l1	0,066	0,123	0,772	0,057	0,046	0,019
Nts	1,686	12,511	0,194	0,607	0,530	0,666
Csf3	0,500	1,456	0,543	0,112	0,000	0,200
Hormad2	0,482	0,289	0,111	0,065	0,015	0,027
Igkv9-123	0,710	2,082	8,834	0,094	0,694	0,561
Ighv1-11	7,371	0,000	0,703	0,665	0,202	0,063
Sftpd	0,166	0,000	10,398	0,064	1,092	0,048
Sfrp4	0,128	0,175	1,120	0,066	0,040	0,049
Ighd	0,178	0,455	1,592	0,091	0,067	0,066
Pck1	26,318	81,057	2,132	5,393	2,263	2,971
Igkv4-90	77,456	897,551	76,291	14,596	21,901	62,710
Mfap5	0,132	1,814	0,732	0,036	0,113	0,074
Slc34a2	0,518	28,348	1,097	0,514	0,524	1,351
Igkv4-61	13,222	253,747	72,396	5,965	11,923	7,834
Naaladl1	6,243	93,131	0,613	1,530	2,156	3,104
Osr2	0,088	17,681	0,337	0,160	0,314	0,650
Ighv9-2	0,477	164,868	7,930	1,731	2,723	6,064
Enpp7	2,123	57,098	0,208	0,415	0,450	2,708
Ighv11-2	0,861	10,959	0,607	0,158	0,000	0,315
Uty	6,719	8,195	6,862	0,049	0,298	0,114
Kdm5d	6,264	5,389	3,798	0,106	0,088	0,044

Table appendix 2.3: Individual expression data from APC P53 and APC P53 RAC1B^{OE} tumours. Gene names and FPKM (fragments kilobase per million reads mapped) values shown individually per each tumour, demonstrating the variability within groups. AP: APC P53; APR: APC P53 RAC1B^{OE}.

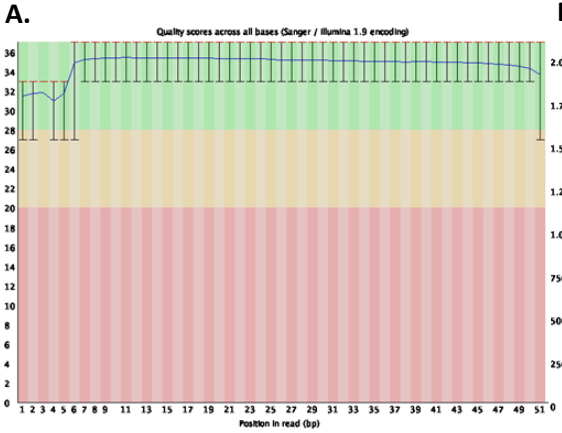
Appendix 2.5: FastQC from RNAseq experiment section 5.2.1.3

Mapping statistics	
APC P53_1	**APC P53 RAC1B^{OE}_1**
reads processed: 37171123	reads processed: 41153699
reads pseudoaligned: 35498422	reads pseudoaligned: 39342936
percentage pseudoaligned: 95.5%	percentage pseudoaligned: 95.6%
GC content (%): 50	GC content (%): 50
Sequence length: 51	Sequence length: 51
APC P53_2	**APC P53 RAC1B^{OE}_2**
reads processed: 40435706	reads processed: 35217377
reads pseudoaligned: 38818278	reads pseudoaligned: 33703030
percentage pseudoaligned: 96.0%	percentage pseudoaligned: 95.7%
GC content (%): 50	GC content (%): 50
Sequence length: 51	Sequence length: 51
APC P53_3	**APC P53 RAC1B^{OE}_3**
reads processed: 31914841	reads processed: 51251914
reads pseudoaligned: 30542503	reads pseudoaligned: 49355593
percentage pseudoaligned: 95.7%	percentage pseudoaligned: 96.3%
GC content (%): 50	GC content (%): 51
Sequence length: 51	Sequence length: 51

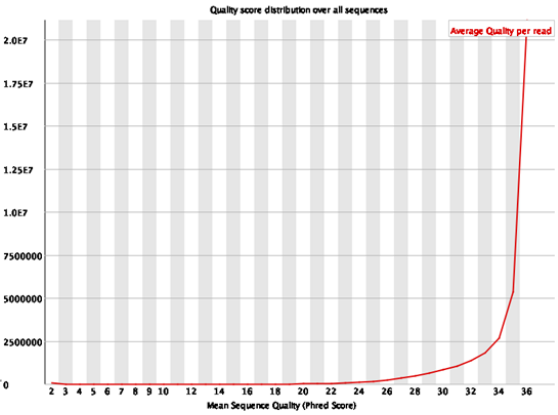
Per base sequence quality (A), per sequence quality scores (B) and per base sequence content (C) modules from the FastQC report are shown below individually per each sample:

APC P53_1:

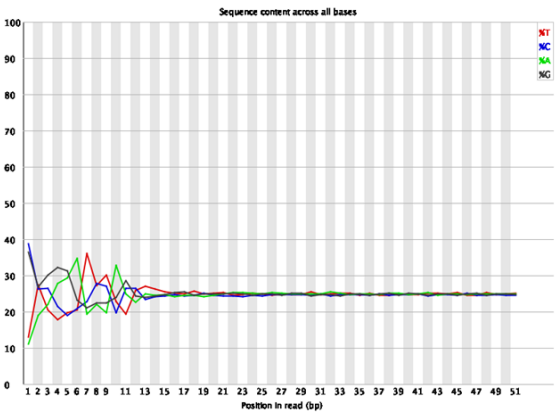
A.



B.

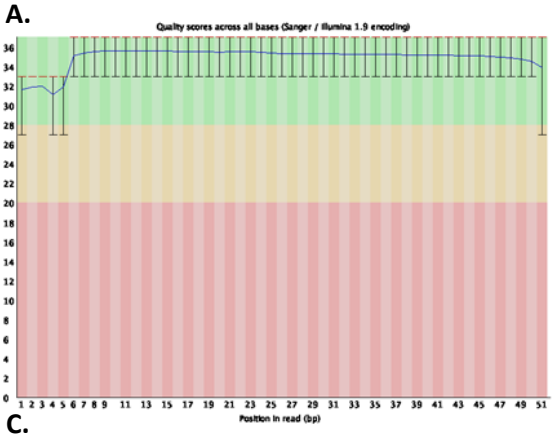


C.

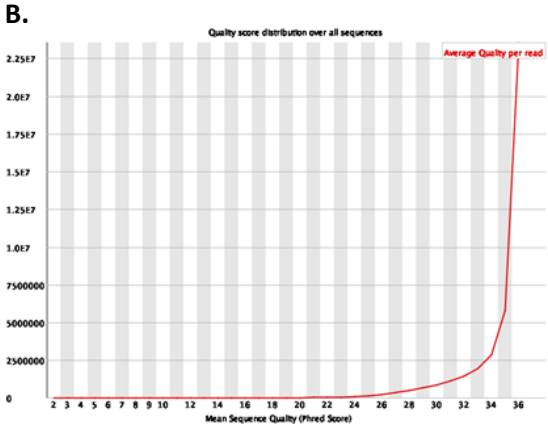


APC P53_2:

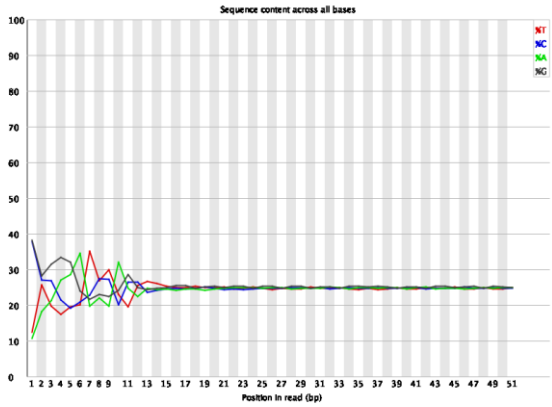
A.



B.

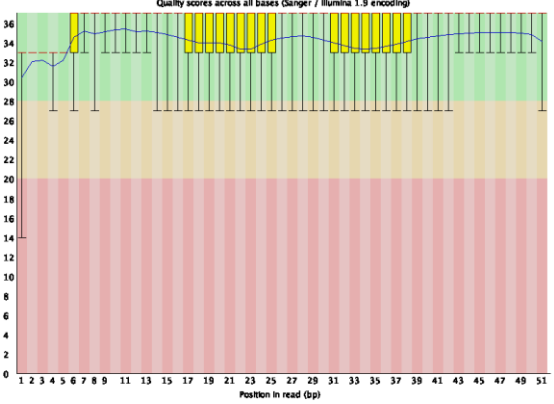


C.

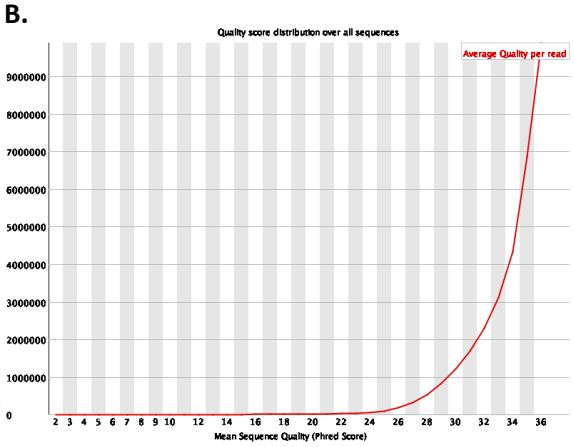


APC P53_3:

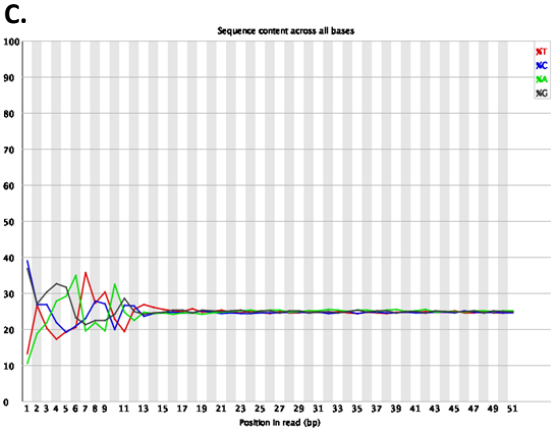
A.



B.

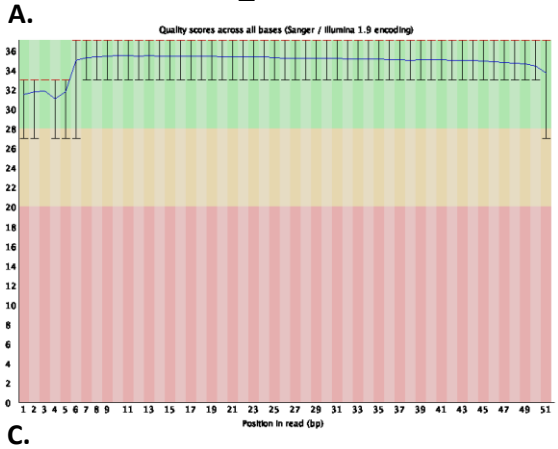


C.

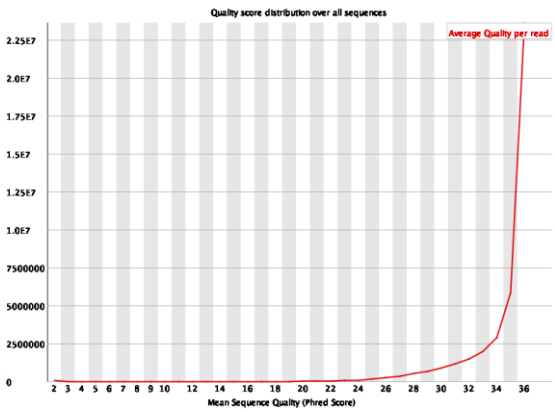


APC P53 RAC1B^{OE}_1:

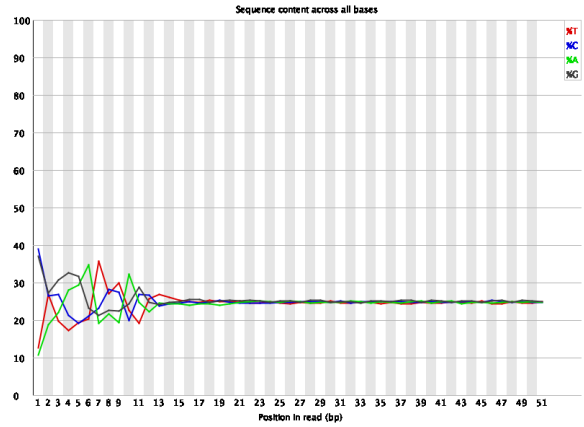
A.



B.

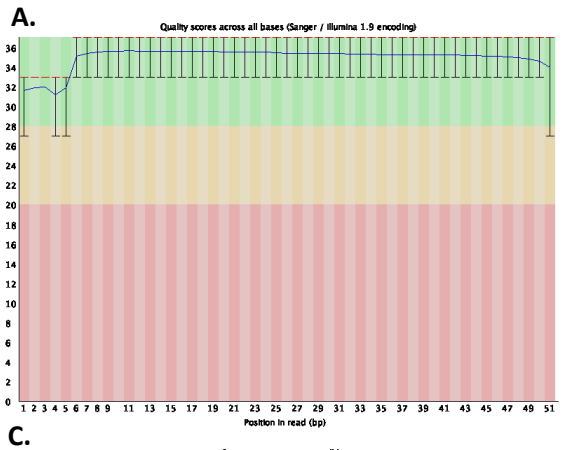


C.

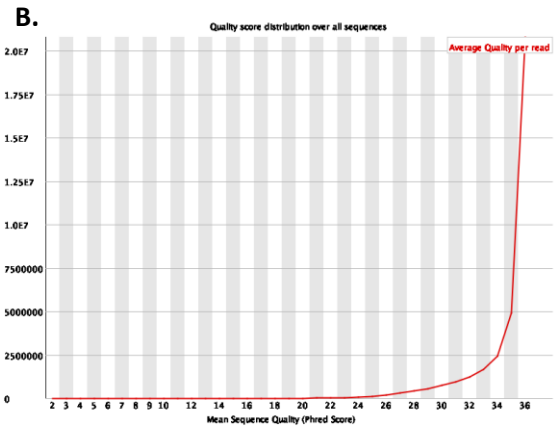


APC P53 RAC1B^{OE}_2:

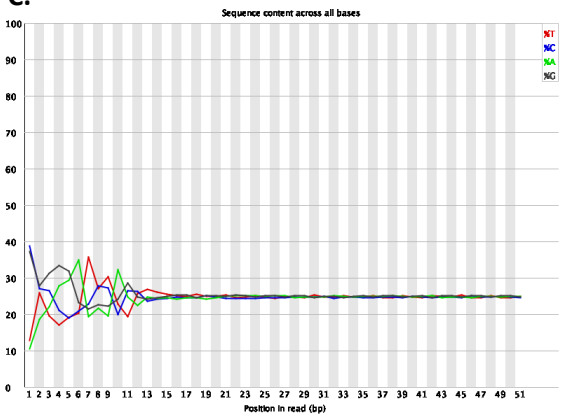
A.



B.

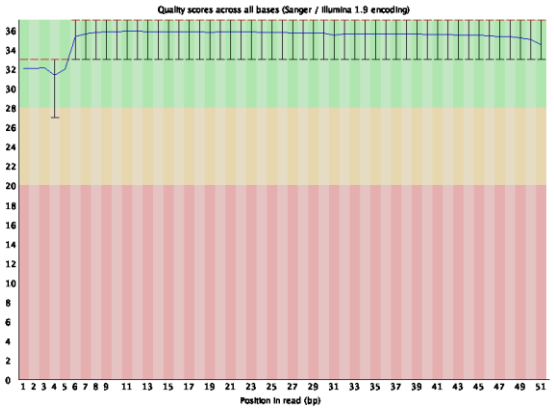


C.

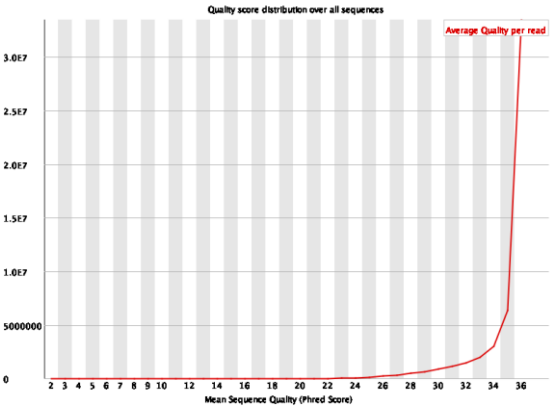


APC P53 RAC1B^{OE}_3:

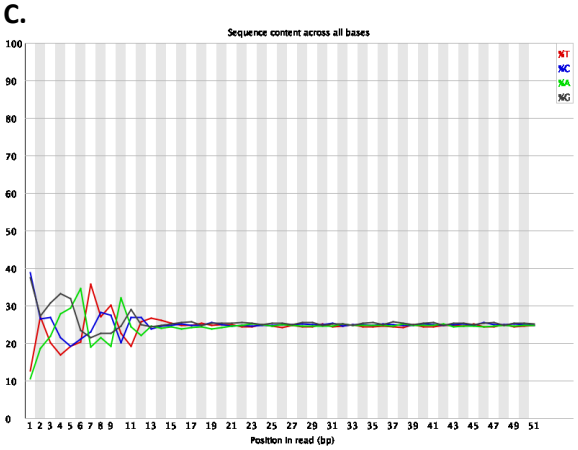
A.



B.



C.



Appendix 3: GSEA list of genes

SANSOM_APC_TARGETS			SANSOM_WNT_PATHWAY_REQUIRE_MYC
AASDH	FKBP5	POLR3D	ACTL6A
ABCE1	FLYWCH2	PRDX2	APOD
ADAT1	FZD6	PRSS2	ARX
ADCY8	GALNT10	PRSS41	ASCL2
AFAP1L1	GGH	PSAT1	AXIN2
AGR3	GLMN	PSRC1	BAX
AKAP13	GPSM1	PVRL3	BMP7
ANAPC4	GULOP	RAD51AP1	CCND2
ANKRD11	HAGHL	RALGAPA1	CCNE2
ANKRD44	HHAT	RASSF4	CD44
AQP4	HIRA	RGS12	CITED1
ARMC9	HMGA2	RNASE1	CSNK2A1
ASCL2	HMGCS2	RNF149	CUL1
ATG16L2	HOPX	RNF43	DVL3
ATG9B	HOXA9	ROR2	E2F1
AXIN2	HS3ST1	RPRD1A	EDN1
BCL11A	HSPBAP1	SESN2	EPHB2
BRD8	HUNK	SHISA2	EPHB3
C10orf107	IFFO2	SHMT2	FOXA1
C13orf15	IGFBP4	SIVA1	FRZB
C16orf88	IKBKAP	SLC1A2	FXN
C2orf88	ILDR1	SLC7A5	FZD6
C9orf3	KIAA0922	SLFN12	FZD7
C9orf41	KIAA1456	SLTM	GEMIN2
CAMTA1	KIAA1524	SMYD5	ID2
CBWD1	KRT23	SNHG12	ID3
CCDC101	KRTAP21-1	SNW1	IGFBP2
CCDC141	LECT2	SNX24	LDB1
CCDC163P	LOC729810	SOX17	LECT2
CCDC86	LPCAT4	SOX4	LEF1
CCDC94	MAD1L1	SOX9	LGR5
CCND2	MAPK4	SP5	MMP14
CD44	MBOAT1	SPACA7	MSX1
CD80	MECOM	SPATA5L1	MTAP
CDC37	MED13L	SRCAP	MYC
CDCA7	MEG3	SRRD	NKD1
CELA2A	METTL13	STOML1	NLK
CELA3B	METTL2B	SYTL1	PARP1
CENPP	MMP15	TBC1D7	PCBD1
CEP72	MND1	TBC1D9	PLAT
CITED1	MRI1	TC2N	PTCH1
CLASP2	MRPL47	TCHP	PTGS2
CLDN11	MRPS14	TGIF2	RABEP2
COX4NB	MSI1	THOC1	ROR2
CPA1	MT4	TIAM1	RUVBL1
CTRB1	MTHFD1L	TMEM209	SEMA3C
CYBRD1	MUTYH	TNFRSF12A	SIM2
DCAF4	MYC	TNFRSF19	SLC1A5
DCLRE1B	MYCL1	TNNI1	SOX17
DCTD	MYL7	TPD52L1	SOX4
DGKB	MYO1B	TRIB1	SP5
DHX30	NDC80	TRIP11	TCF7
DIS3L	NEB	TRIT1	TIAM1
DMP1	NFS1	TRMT1	TNFRSF12A
DNMT1	NINL	TSEN2	TNFRSF19
DTL	NR2E3	TTBK2	TP53
DUSP4	NT5C3L	TTC27	WIF1
DYRK3	OBFC2A	TUBB2B	WISP1
EDN1	ODZ4	TXNRD3	
EIF5B	ORC5	TYMS	
EMP2	PAICS	VNN1	
EPHB2	PASK	WASF3	
EPHB3	PHF21B	WDR34	
EPN3	PIEZO1	WDR62	
ERCC4	PLA2G15	ZFP1	
ETS2	PLA2G2F	ZNF184	
ETV4	PLCD3	ZNF627	
FADS3	PMM1	ZNF746	
FAM20B	PNLIPRP1	ZNF800	
FBLN2	POGK	ZSCAN21	
FGFRL1	POLE		

Table continues in the next page.

Lgr5 high vs low study C		Direct and functional βcatenin targets in SW480		
4933406C10RIK	LDHB	ABHD12B	FRAS1	NOTUM
9930013L23RIK	LECT2	ABHD6	FYN	NPNT
A730054J21RIK	LGR5	ACSS2	GABARAPL1	NR2F1
A930009A15RIK	LIMCH1	AMDHD1	GLI3	NUDT4
ABTB2	MARVELD1	ANKRD1	GPSM2	OXGR1
AFAP1L1	MDFIC	ANKRD38	GSDM1	PAG1
AGR3	MEX3A	ANKRD57	GSTA4	PBX4
AKR1D1	MNX1	APCDD1	GYLTL1B	PCMTD2
ANGPT2	MPP3	ARFIP1	HSDL1	PCP4
APCDD1	NAV1	ASPSCR1	HUNK	PIM1
ASCL2	NOTCH4	ATP12A	HYAL1	PKIG
ASGR1	NR2E3	AXIN2	ID2	PKP2
ASGR2	NRP1	BMP7	IFI27	PLEKHO1
AXIN2	ODZ4	C10ORF118	IL10	PPP1R2
B230114P17RIK	PCDH8	C12ORF34	IL10RA	PRMT3
BCL2	PGC	C12ORF4	IL17RD	PROX1
BEX1	PIK3IP1	C16ORF30	KCNQ1DN	PSTPIP2
BEX4	PLP1	C20ORF118	KIAA0888	PVALB
BHLHB9	PTGDS	C20ORF19	KIAA0895	RAB11FIP4
BRAF	PXDN	C3ORF21	KIAA1199	RASL10B
CAP2	RASSF4	C6ORF130	KIAA1274	RASL11B
CBS	RASSF5	C6ORF97	KLRG1	RTKN
CCDC3	RGMB	C7ORF52	KRT20	S100A3
CCDC46	RGS19	CCDC84	KRT23	SASP
CCND3	RGS7BP	CDX2	KRT71	SDK1
CHST11	RPE	CEP78	KRT74	38231
CHST11 /// PHACTR1	SCN2B	CHRNA3	LEF1	SLC16A1
CHST2	SECTM1A	CHST13	LGR6	SLC7A6
CIB2	SEMA3C	CKMT2	LHX4	SLC7A8
CYP2E1	SERTAD4	CPXM1	LINCR	SMAD7
D5ERTD505E	SLC1A2	CRABP2	LMBR1	SMTNL2
DGKG	SLC25A27	CRYGN	LOC145757	SNX25
DLGAP1	SLC38A4	CST9	LOC157860	SORBS2
DYNC2H1	SLCO3A1	CTR9	LOC286002	SP5
E230029C05RIK	SLIT2	CTSZ	LOC340073	STX3
EPHA4	SMARCD3	CXCL14	LOC644662	SYK
ESRRG	SMO	CXCR4	LRIG1	TDGF1
EXPI	SOAT1	CXCR6	LRRC2	THSD3
FGFR4	SORBS2	DDX26B	MAP3K8	TMEM47
FMNL2	SORCS2	DEPDC7	MFSD11	TNFRSF11B
FSCN1	SOX4	DLX2	MSX1	TNFRSF19
FSTL1	SRPK2	DLX4	MTSS1	TNFRSF1B
GPER	TACC1	DOCK11	MYH7B	TNIK
GPR126	TMEM146	EDAR	MYO1D	TNNC1
GRAMD1A	TMEM182	EFNA4	MYO1F	TREM2
GZMA	TNFRSF19	EN2	NAPG	TTC29
HNMT	TRF	ENTPD8	NEK9	UBXD3
ID2	VAV3	ETV3	NELF	UGT8
IGFBP4	VWA2	FASLG	NKD1	VEPH1
IL17RD	WFDC15B	FGF19	NKD2	WNT11
IRS1	WWTR1	FGF3	NKPD1	WNT6
KIF12	ZBTB16	FGF9	NLRP11	YAF2
KIF26B	ZFP383	FLJ11710	NOLC1	ZAP70
KLHL13	ZFP618	FLJ35767	NOS2A	ZNRF3

Table appendix 3: List of genes within each dataset utilised for GSEA. The "SANSOM_APC_TARGETS" list are genes up-regulated after Cre-lox knockout of APC (GeneID=324) in the small intestine. The "SANSOM_WNT_PATHWAY_REQUIRE_MYC" list are Wnt target genes up-regulated after Cre-lox knockout of APC (GeneID=324) in the small intestine that require functional MYC (GeneID=4609). The "Lrg5 high vs low study C" list is derived from the supplementary data (Table S1) of the Merlos-Suarez *et al.* paper from 2011 (murine Lgr5 refined ISC signature). The "Direct and functional βcatenin target genes in SW480" is derived from the supplementay data (Table S1) of the Watanabe *et al*. paper from 2014 (combined analysis of βcatenin target genes in SW480 cancer cells by CHIPseq and siRNA depletion.

Appendix 4: <u>BioID protein list</u>

Protein IDs	Protein names	Gene names	Sequence coverage [%]	Mol. weight [kDa]	Sequence length	MS/MS count _BirA1	MS/MS count _BirA_2	MS/MS count _BirA_3	MS/MS count _Rac1_1	MS/MS count _Rac1_2	MS/MS count _Rac1_3	MS/MS count _Rac1b_1	MS/MS count _Rac1b_2	MS/MS count _Rac1b_3	MS/MS count
P63001;Q051 44;P60764	Ras-related C3 botulinum toxin substrate 1;Ras-related C3 botulinum toxin substrate 2	Rac1;Rac2	35,9	21,45	192	7	3	2	37	29	28	20	25	16	167
Q62470	Integrin alpha-3;Integrin alpha-3 heavy chain;Integrin alpha-3 light chain	Itga3	21,3	116,74	1053	10	11	10	26	22	26	23	29	34	191
Q01279;Q615 27	Epidermal growth factor receptor	Egfr	18,1	134,85	1210	7	3	5	24	16	20	16	17	14	122
Q60634	Flotillin-2	Flot2	35,3	47,037	428	4	1	3	12	12	10	8	10	9	69
Q9DBJ3	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1	Baiap2l1	37	57,187	514	2	3	3	18	10	11	14	11	13	85
Q6P9J9	Anoctamin-6	Ano6	13,1	106,25	911	4	5	6	12	5	7	6	9	9	63
Q91W92	Cdc42 effector protein 1	Cdc42ep1	29,1	43,095	409	2	1	0	11	10	7	10	9	5	55
Q9WVL3	Solute carrier family 12 member 7	Slc12a7	22,5	119,48	1083	3	0	1	21	13	13	18	10	14	93
P97792	Coxsackievirus and adenovirus receptor homolog	Cxadr	25,2	39,947	365	5	8	4	11	5	8	7	6	8	62
O54990	Prominin-1	Prom1	18,7	97,112	867	5	6	4	16	11	10	10	8	12	82
Q8BKX1	Brain-specific angiogenesis inhibitor 1-associated protein 2	Baiap2	44,1	59,236	535	0	1	0	12	10	11	14	8	6	62
Q8CGB3	Uveal autoantigen with coiled-coil domains and ankyrin repeats	Uaca	19,6	160,81	1411	1	0	0	7	18	17	9	14	8	74
O09044	Synaptosomal-associated protein 23	Snap23	46,2	23,261	210	2	4	4	11	6	9	5	7	7	55
E9Q912	Transmembrane protein 237	Rap1gds1	36,4	66,076	607	3	1	4	7	5	7	24	18	18	87
Q8K078	Solute carrier organic anion transporter family member 4A1	Slco4a1	7,3	77,668	723	1	1	1	5	3	4	3	5	2	25
Q9QWY8	Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 1	Asap1	12,8	127,42	1147	3	1	0	10	6	9	6	4	3	42
Q64455	Receptor-type tyrosine-protein phosphatase eta	Ptprj	10,5	136,77	1238	0	0	0	7	8	5	2	2	3	27
Q60738	Zinc transporter 1	Slc30a1	9,7	54,716	503	3	0	1	5	2	4	1	2	3	21
A2ATK9		Fam171a1	12,9	97,961	892	0	0	0	3	2	4	3	1	4	17
Q8BH43	Wiskott-Aldrich syndrome protein family member 2	Wasf2	6,8	54,073	497	1	1	2	2	2	4	2	2	3	17
Q3V0J1	Transmembrane protein 237	Tmem237	27,6	47,342	427	0	0	0	8	4	4	4	5	4	29
O55098	Serine/threonine-protein kinase 10	Stk10	12,7	111,9	966	1	0	0	3	3	5	2	3	4	21
Q8VDZ4	Palmitoyltransferase ZDHHC5	Zdhhc5	11,9	77,5	715	0	0	0	5	3	3	3	2	2	18
Q8BGT1		Flrt3	14,2	72,878	649	0	0	0	3	1	3	2	2	3	14
P28571	Sodium- and chloride-dependent glycine transporter 1	Slc6a9	8,2	76,543	692	1	0	1	2	1	2	1	1	1	10
Q5FWI3	Transmembrane protein 2	Tmem2	6,9	153,8	1383	0	0	0	5	2	2	2	3	0	14
Q5U431	G-protein coupled receptor 39	Gpr39	6,8	51,589	456	0	0	1	2	3	2	2	1	3	14
Q61146	Occludin	Ocln	7,3	58,999	521	0	0	0	1	2	0	1	0	0	4
F7BIK1		Pcdh1	3,1	116,52	1077	0	0	0	1	0	2	1	0	1	5
Q6PD21	SH2 domain-containing adapter protein B	Shb	7,4	54,707	503	0	0	0	2	3	2	1	1	1	10
P70424	Receptor tyrosine-protein kinase erbB-2	Erbb2	6,1	138,58	1256	0	0	0	4	2	2	2	0	0	10
Q62469	Integrin alpha-2	Itga2	3,5	128,95	1178	1	0	0	2	2	2	2	1	1	11
O70309	Integrin beta-5	Itgb5	7,4	87,908	798	0	0	1	2	1	2	2	2	2	12
P52800	Ephrin-B2	Efnb2	22	37,202	336	0	0	0	3	3	5	0	1	2	14
P0C673	Immunoglobulin superfamily member 11	Igsf11	2,8	46,067	428	0	0	0	1	1	1	1	1	1	6
Q80W68	Kin of IRRE-like protein 1	Kirrel	6,7	87,175	789	0	0	0	2	3	3	1	0	0	9
B5TVM2	Immunoglobulin-like domain-containing receptor 2	Ildr2	10	73,235	661	0	0	0	5	1	2	1	0	1	10
Q9JIM1	Equilibrative nucleoside transporter 1	Slc29a1	6,1	50,192	460	0	0	0	2	2	2	2	1	1	10
P97798	Neogenin	Neo1	4,6	163,16	1493	0	0	0	3	2	2	1	1	2	11
Q8BTW9	Serine/threonine-protein kinase PAK 4	Pak4	3,7	64,622	593	0	1	1	1	1	2	1	1	2	10
Q69Z29	Uncharacterized protein KIAA0754	Kiaa0754	2,2	104,64	979	0	0	1	1	1	2	0	1	1	8
Q8R3T5	Syntaxin-binding protein 6	Stxbp6	10	23,671	210	0	0	0	2	1	1	0	0	0	4
Q04736	Tyrosine-protein kinase Yes	Yes1	9,8	60,629	541	0	0	0	1	0	2	0	0	0	3
P54818	Galactocerebrosidase	Galc	5,7	77,256	684	0	0	1	1	2	2	0	0	1	7
Q8BGU5	Cyclin-Y	Ccny	4,4	39,394	341	1	0	0	1	1	1	1	1	0	6
P59281	Rho GTPase-activating protein 39	Arhgap39	2,3	125,2	1107	0	0	0	2	2	1	0	0	0	5
Q9CQX5	Claudin domain-containing protein 1	Cldnd1	5,9	28,572	253	0	0	0	1	1	1	1	0	1	5
P20352	Tissue factor	F3	8,2	32,935	294	0	0	0	1	1	1	0	0	0	3
Q9CQS5	Serine/threonine-protein kinase RIO2	Riok2	10,6	62,49	547	0	0	1	0	0	0	0	2	0	3
Q9JI59	Junctional adhesion molecule B	Jam2	7	33,047	298	0	0	0	0	1	1	0	1	0	3
A4Q9E5	Tubulin monoglycylase TTLL3	Ttll3	1,6	104,43	927	0	0	0	0	0	0	1	1	0	2
REV__Q14B62			0	101,5	888	0	0	0	0	1	0	0	0	0	1
Q3U2A8	Valine--tRNA ligase, mitochondrial	Vars2	2,5	118,46	1060	0	0	0	0	2	1	2	0	0	5
P97355	Spermine synthase	Sms	8,2	41,313	366	0	0	0	1	0	1	0	0	0	2
Q3TGW2	Endonuclease/exonuclease/phosphatase family domain-containing protein 1	Eepd1	2,6	62,953	569	0	0	0	1	0	0	0	0	0	1
Q62504	Msx2-interacting protein	Spen	0,4	398,75	3644	0	0	0	0	0	0	0	1	1	2
Q60751	Insulin-like growth factor 1 receptor;Insulin-like growth factor 1 receptor alpha chain;Insulin-like growth factor 1 receptor beta chain	Igf1r	1,7	155,79	1373	0	0	0	1	1	1	0	0	2	5
REV__M0QWR8			0	73,818	670	0	0	0	0	1	0	0	0	0	1
P47802	Metaxin-1	Mtx1	4,4	35,623	317	1	0	0	1	0	1	2	1	2	8
Q08642	Protein-arginine deiminase type-2	Padi2	2,5	76,249	673	0	0	0	0	0	1	1	1	1	4
Q8R2Q8	Bone marrow stromal antigen 2	Bst2	11	19,152	172	0	0	0	1	0	0	1	0	0	2
REV__Q3V0C5			0	120,63	1052	0	0	0	0	0	0	0	0	1	1
P31725	Protein S100-A9	S100a9	9,7	13,049	113	0	0	0	0	0	0	2	0	0	2
Q3TC33	Coiled-coil domain-containing protein 127	Ccdc127	9,2	30,509	260	0	0	0	0	0	0	0	1	0	1

<u>Table appendix 4</u>: Extended BioID protein list. Protein IDs, protein names and genes names from the significantly different BioID hits. It is also shown the percentage of sequence coverage, the molecular weight and length of proteins and the MS counts per condition and total MS counts per protein.

Source	Num. reference	URL	Date accessed
Cancer Research UK	2	https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer.	24th August 2019
American Institute for Cancer Research	8	https://www.wcrf.org/dietandcancer/cancers	16th July 2019
International Agency for Research on Cancer	22	http://gco.iarc.fr/today	24th August 2019
CBioPortal	183	https://www.cbioportal.org	27th May 2019
UCSC Xena	451	http://xena.ucsc.edu/?ref=labworm.	27th May 2019
TCGA SpliceSeq	441	http://projects.insilico.us.com/TCGASpliceSeq/index.jsp	16th May 2019
TSVdb	453	http://www.tsvdb.com/	14th May 2019
The Cancer Proteome Atlas	454	https://tcpaportal.org/tcpa/download.html	6th June 2019
Innate DB		https://www.innatedb.com/	5th April 2019
Synapse	440	https://www.synapse.org/#!Synapse:syn300013;https://www.synapse.org/#!Synapse:syn2623706	7th May 2019

Table appendix 5: Pages accessed. List of data sources used for this thesis, annotated with their URLs, data of access and number of reference according to bibliography.